(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 July 2001 (26.07.2001)

PCT

(10) International Publication Number
WO 01/53344 A2

(51) International Patent Classification[7]: C07K 14/47

(21) International Application Number: PCT/US01/01700

(22) International Filing Date: 18 January 2001 (18.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/176,898 19 January 2000 (19.01.2000) US
09/724,310 28 November 2000 (28.11.2000) US

(71) Applicant *(for all designated States except US)*: **AMGEN, INC.** [US/US]; One Amgen Center Drive, Thousand Oaks, CA 91320-1799 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **NGUYEN, Hung** [US/US]; 881 St. Charles Drive, Apt. 10, Thousand Oaks, CA 91360 (US). **JUNG, Jeannie** [US/US]; Apt. 227, 5348 Newcastle Avenue, Encino, CA 91316 (US). **CLARKIN, Kristie** [US/US]; 2318 Eagle Creek Lane, Oxnard, CA 93030 (US).

(74) Agent: **CLOUGH, David, W.**; Marshall, O'Toole, Gerstein, Murray & Borun, 6300 Sears Tower, 223 South Wacker Drive, Chicago, IL 60606 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/53344 A2

(54) Title: CHONDROMODULIN-I RELATED PEPTIDE

(57) Abstract: The present invention relates to a novel polynucleotide encoding a polypeptide which is structurally related to chondromodulin-I, as well as to the purified protein itself. The invention also relates to vectors, host cells, antibodies and recombinant methods for producing the polypeptide. In addition, the invention discloses therapeutic, diagnostic and research utilities for these and related products.

CHONDROMODULIN-I RELATED PEPTIDE

RELATED APPLICATION

This application claims priority from United States application 09/724,310 filed November 28, 2000 and United Stated provisional application 60/176,898 filed January 19, 2000.

FIELD OF INVENTION

The present invention relates to a novel polypeptide that is related to chondromodulin-I, referred to as ChMIrp, and nucleic acid molecules encoding the polypeptide. The invention also relates to vectors, host cells, selective binding agents such as antibodies, and methods of producing ChMIrp polypeptides. Also provided for are methods for the use of ChMIrp, including methods for the diagnosis and treatment of disorders associated with ChMIrp.

BACKGROUND

Technical advances in the identification, cloning, expression and manipulation of nucleic acid molecules have greatly accelerated the discovery of novel therapeutics based upon deciphering the human genome. Rapid nucleic acid sequencing techniques can now generate sequence information at unprecedented rates and, coupled with computational analyses, allow the assembly of overlapping sequences into the entire genome and the identification of polypeptide-encoding regions. Comparison of a predicted amino acid sequence against a database compilation of known amino acid sequences can allow one to determine the extent of homology to previously identified sequence and/or structure landmarks. Cloning and expression of a polypeptide-encoding region of a nucleic acid molecule provides a polypeptide product for structural and functional analysis. Manipulation of nucleic acid molecules and encoded polypeptides to produce variants and derivatives thereof may confer advantageous properties on a product for use as a therapeutic.

In spite of the significant technical advances in genome research over the past decade, the potential for development of novel therapeutics based on the human genome is still largely unrealized. While a number of genes encoding potentially beneficial protein therapeutics, or those encoding polypeptides which may act as "targets" for therapeutic molecules, have been identified using recombinant DNA technology, the structure and function of a vast number of genes in the genome of mammals are yet unknown.

Accordingly, it is an object of the invention to identify novel polypeptides and nucleic acid molecules encoding the same which have diagnostic or therapeutic benefit.

Chondromodulin-I (ChM-I) was first identified as a proteinous component within fetal bovine cartilage extracts that in the presence of fibroblast growth factor-2 (FGF-2) stimulated DNA synthesis of rabbit cultured growth plate chondrocytes. Subsequently, this growth stimulatory factor was isolated and the amino acid sequence determined. Degenerate primers, based on the amino acid sequence, were used to PCR clone a fragment of the bovine gene. This DNA fragment was used to identify a 1.7 kb band in bovine epiphysial cartilage mRNA and was then used to isolate the bovine chondromodulin-I gene from a bovine epiphysial cartilage cDNA library (Hiraki *et al.*, *Biochem. Biophys. Res. Com.*, 175: 971-974, 1999). Subsequently, ChM-I orthologs from human, mouse, rat, rabbit and chicken have been isolated; *see* Hiraki *et al.* (*Eur. J. Biochemistry*, 260: 869-878, 1999) and Shukunami and Hiraki (*Biochem. Biophys. Res. Com.*, 249: 885-890, 1998) for the isolation of the human and rabbit orthologs, respectively. ChM-I has been found to be expressed by chondrocytes (Hiraki *et al.*, *Eur. J. Biochemistry*, 260: 869-878, 1999).

Expression of the human ChM-I gene in CHO cells revealed that the secreted mature protein was larger than the polypeptide isolated from cartilage extracts (Hiraki *et al.*, *Eur. J. Biochemistry*, 260: 869-878,1999). This proteolytic processing was also demonstrated when the ChM-I rabbit ortholog was expressed in monkey COS cells (Shukunami and Hiraki, *Biochem. Biophys. Res. Com.*, 249: 885-890, 1998). Alignment of the amino acid sequences of the mature and

precursor proteins revealed that the mature protein sequence begins immediately after the RERR amino acid sequence present in the precursor form. The ChM-I precursor contains a single hydrophobic region near its amino-terminal which is thought to be involved in membrane insertion. The RERR sequence acts as a processing signal that mediates proteolytic cleavage which results in the secretion of the mature form. This processing causes the majority of the amino-terminal portion of the protein to remain inserted within the cellular membrane of chondrocytes. Thus, these results indicate that ChM-I is expressed as a larger transmembrane precursor protein that is cleaved into the carboxy-terminal mature form and deposited into cartilage (reviewed in Suzuki, *Biochem. Biophys. Res. Comm.* 259: 1-7, 1999).

Expression of ChM-I mRNA has been detected only in the cartilage of human and bovine developing embryos. Specifically, high levels of expression were detected in chondrocytes residing in the proliferating cartilage zone with lower levels detected in chondrocytes residing in the adjacent resting and upper hypertrophic zones. No discernible expression was detected in chondrocytes residing in the articular and lower calcified hypertrophic zones. ChM-I polypeptide was localized to the inter-territorial region of the proliferating, resting and upper hypertrophic zones (Hiraki *et al.*, *Eur. J. Biochemistry*, 260: 869-878, 1999).

ChM-I polypeptide can stimulate DNA and proteoglycan synthesis in rabbit culture growth plate chondrocytes in the presence or absence of FGF-2. ChM-I inhibits the growth and tube morphogenesis in bovine carotid artery endothelial cells, *in vitro*, and fine capillary formation in a chicken chorioallantoic membrane assay, *in vivo* (Hiraki *et al., Eur. J. Biochemistry*, 260: 869-878, 1999). ChM-I polypeptide synergises with FGF-2 to induce soft-agar colony formation of cultured growth plate chondrocytes (Inoue *et al.*, *Biochem. Biophys. Res. Com.*, 241: 395-400, 1999). Primary osteoblasts and MC3T3-e1 osteoblast-like cells proliferate when stimulated with ChM-I (Mori *et al.*, *FEBS Lett.*, 406: 310-314, 1999).

Thus, identification of chondromodulin-I has led to a better understanding of the processes involved in mediating the development of skeletal

components, including bone growth, cartilage formation and inhibition of cartilage vascularization. Identification of the chondromodulin-I related gene and polypeptide as described herein and other chondromodulin related polypeptides will further clarify the understanding of these processes and facilitate the development of therapies for pathological conditions which involve the degradation of skeletal components and increased vascularization.

SUMMARY OF INVENTION

The present invention relates to a novel serine/threonine kinase family and uses thereof. More specifically, the present invention relates to novel ChMIrp nucleic acid molecules and encoded polypeptides, and uses thereof.

The invention provides for an isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

(a) the nucleotide sequence set forth in SEQ ID NO: 1;

(b) a nucleotide sequence encoding the polypeptide set forth in SEQ ID NO: 2;

(c) a nucleotide sequence which hybridizes under moderately or highly stringent conditions to the complement of (a) or (b), wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2; and

(d) a nucleotide sequence complementary to any of (a) through (c).

The invention also provides for an isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence encoding a polypeptide that is at least about 70, 75, 80, 85, 90, 95, 96, 97, 98, or 99 percent identical to the polypeptide set forth in SEQ ID NO: 2, wherein the polypeptide has an activity of the encoded polypeptide set forth in SEQ ID NO: 2 as determined using a computer program selected from the group consisting of GAP, BLASTP, BLASTN, FASTA, BLASTA, BLASTX, BestFit, and the Smith-Waterman algorithm;

(b) a nucleotide sequence encoding an allelic variant or splice variant of the nucleotide sequence set forth in SEQ ID NO: 1, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(c) a nucleotide sequence of SEQ ID NO: 1, (a), or (b) encoding a polypeptide fragment of at least about 25 amino acid residues, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(d) a nucleotide sequence encoding a polypeptide that has a substitution and/or deletion of 1 to 250 amino acid residues set forth in any of SEQ ID NOS: 1-2 wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(e) a nucleotide sequence of SEQ ID NO: 1, or (a)-(d) comprising a fragment of at least about 16 nucleotides;

(f) a nucleotide sequence which hybridizes under moderately or highly stringent conditions to the complement of any of (a)-(e), wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2; and

(g) a nucleotide sequence complementary to any of (a)-(e).

The invention further provides for an isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence encoding a polypeptide set forth in SEQ ID NO: 2 with at least one conservative amino acid substitution, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(b) a nucleotide sequence encoding a polypeptide set forth in SEQ ID NO: 2 with at least one amino acid insertion, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(c) a nucleotide sequence encoding a polypeptide set forth in SEQ ID NO: 2 with at least one amino acid deletion, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(d) a nucleotide sequence encoding a polypeptide set forth in SEQ ID NO: 2 which has a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(e) a nucleotide sequence encoding a polypeptide set forth in SEQ ID NO: 2 with at least one modification selected from the group consisting of amino acid substitutions, amino acid insertions, amino acid deletions, C-terminal truncation, and N-terminal truncation, wherein the polypeptide has an activity of the encoded polypeptide set forth in SEQ ID NO: 2;

(f) a nucleotide sequence of (a)-(e) comprising a fragment of at least about 16 nucleotides;

(g) a nucleotide sequence which hybridizes under moderately or highly stringent conditions to the complement of any of (a)-(f), wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2; and

(h) a nucleotide sequence complementary to any of (a)-(e).

The invention also provides for an isolated polypeptide comprising the amino acid sequence selected from the group consisting of:

(a) the mature amino acid sequence set forth in SEQ ID NO: 2 comprising a mature amino terminus at residue 1, and optionally further comprising an amino-terminal methionine;

(b) an amino acid sequence for an ortholog of SEQ ID NO: 2, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(c) an amino acid sequence that is at least about 70, 75, 80, 85, 90, 95, 96, 97, 98, or 99 percent identical to the amino acid sequence of SEQ ID NO: 2, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2 as determined using a computer program selected from the group consisting of GAP, BLASTP, BLASTN, FASTA, BLASTA, BLASTX, BestFit, and the Smith-Waterman algorithm;

(d) a fragment of the amino acid sequence set forth in SEQ ID NO: 2 comprising at least about 25 amino acid residues, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(e) an amino acid sequence for an allelic variant or splice variant of either the amino acid sequence set forth in SEQ ID NO: 2, or at least one of (a)-(c) wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2.

The invention further provides for an isolated polypeptide comprising the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth in SEQ ID NO: 2 with at least one conservative amino acid substitution, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(b) the amino acid sequence set forth in SEQ ID NO: 2 with at least one amino acid insertion, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(c) the amino acid sequence set forth in SEQ ID NO: 2 with at least one amino acid deletion, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(d) the amino acid sequence set forth in SEQ ID NO: 2 which has a C- and/or N-terminal truncation, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2; and

(e) the amino acid sequence set forth in SEQ ID NO: 2, with at least one modification selected from the group consisting of amino acid substitutions, amino acid insertions, amino acid deletions, C-terminal truncation, and N-terminal truncation, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2.

Also provided are fusion polypeptides comprising the polypeptide sequences of (a)-(e) above of the preceding paragraphs.

The present invention also provides for an expression vector comprising the isolated nucleic acid molecules set forth herein , recombinant host cells comprising recombinant nucleic acid molecules set forth herein, and a method of producing a ChMIrp polypeptide comprising culturing the host cells and optionally isolating the polypeptide so produced.

A transgenic non-human animal comprising a nucleic acid molecule encoding a ChMIrp polypeptide is also encompassed by the invention. The ChMIrp nucleic acid molecules are introduced into the animal in a manner that allows expression and increased levels of the ChMIrp polypeptide, which may include

increased circulating levels. The transgenic non-human animal is preferably a mammal.

Also provided are derivatives of the ChMIrp polypeptides of the present invention.

Analogs of ChMIrp are provided for in the present invention which result from conservative and non-conservative amino acids substitutions of the ChMIrp polypeptide of SEQ ID NO: 2. Such analogs include a ChMIrp polypeptide wherein the amino acid at position 276 is selected from the group consisting of cysteine, serine or alanine, a ChMIrp polypeptide wherein the amino acid at position 280 is selected from the group consisting of cysteine, serine or alanine, a ChMIrp polypeptide wherein the amino acid at position 281 is selected from the group consisting of glutamic acid or aspartic acid, a ChMIrp polypeptide wherein the amino acid at position 285 is selected from the group consisting of glycine, proline or alanine, a ChMIrp polypeptide wherein the amino acid at position 297 is selected from the group consisting of arginine, lysine, glutamine or asparagine, a ChMIrp polypeptide wherein the amino acid at position 300 is selected from the group consisting of cysteine, serine or alanine, a ChMIrp polypeptide wherein the amino acid at position 306 is selected from the group consisting of cysteine, serine or alanine, and a ChMIrp polypeptide wherein the amino acid at position 310 is selected from the group consisting of valine, isoleucine, methionine, leucine, phenylalanine, alanine or norleucine.

Additionally provided are selective binding agents such as antibodies and peptides capable of specifically binding the ChMIrp polypeptides of the invention. Such antibodies and peptides may be agonistic or antagonistic.

Pharmaceutical compositions comprising the nucleotides, polypeptides, or selective binding agents of the present invention and one or more pharmaceutically acceptable formulation agents are also encompassed by the invention. The pharmaceutical compositions are used to provide therapeutically effective amounts of the nucleotides or polypeptides of the present invention. The invention is also directed to methods of using the polypeptides, nucleic acid

molecules, and selective binding agents. The invention also provides for devices to administer a ChMIrp polypeptide encapsulated in a membrane.

The ChMIrp polypeptides of the invention and their biologically active variants, analogs, homologs and fragments may be used for therapeutic and/or diagnostic purposes to treat, prevent and/or detect conditions resulting from abnormal levels of ChMIrp polypeptide or a susceptibility to pathological conditions involving the overreaction of the host to chondromodulin family members or deficiency of the autoregulatory network controlled by these proteins as frequently observed in skeletal conditions such as osteoporosis and dwarfism, and pathological conditions involving angiogenesis including cancer, inflammatory diseases such as psoriasis and cirrhosis, and vascular diseases such as atherosclerosis, coronary heart disease and hypertension.

The invention provides for treating, preventing or ameliorating diseases resulting from abnormal levels of ChMIrp by administering to a mammal a biologically active ChMIrp polypeptide and/or one or more of its biologically active variants, fragments, homolog or variant, either alone or in conjunction with other therapeutic agents. The invention also provides for a method of diagnosing such disorders/diseases or a susceptibility to such disorders/diseases in an animal comprising using a ChMIrp polypeptide and/or antibody and/or amount of ChMIrp genes in body fluids and/or tissues. The invention also provides for methods of diagnosing disorders/diseases resulting from mutation of the ChMIrp gene using the nucleic acid molecule of the invention as a probe. The animal is preferably mammal and more preferably human.

The invention also provides methods of testing the impact of molecules on the expression of ChMIrp polypeptide. A method of regulating expression and modulating (*i.e.* increasing or decreasing) levels of ChMIrp polypeptides are also encompassed by the invention. One method comprises administering to an animal a nucleic acid molecule *in vivo*. In another method, a nucleic acid molecule comprising elements that regulate expression of a ChMIrp polypeptide may be administered to the animal. Examples of these methods include gene therapy and antisense therapy.

Methods of regulating expression and modulating (*i.e.*, increasing or decreasing) levels of a ChMIrp polypeptide are also encompassed by the invention. One method comprises administering to an animal a nucleic acid molecule encoding a ChMIrp polypeptide. In another method, a nucleic acid molecule comprising elements that regulate or modulate the expression of a ChMIrp polypeptide may be administered. Examples of these methods include gene therapy, cell therapy, and anti-sense therapy as further described herein.

The ChMIrp polypeptide was highly expressed in a wide range of primary human tumors. Therefore, the present polypeptide, and its useful nucleic acid intermediates, have demonstrated utility in differentiating transformed cells from the background.

In another aspect of the present invention, the ChMIrp polypeptides may be used for identifying receptors or binding partners thereof ("ChMIrp receptors" or "ChMIrp binding partners"). Various forms of "expression cloning" have been extensively used to clone receptors for protein or co-factors. *See,* for example, Simonsen and Lodish, *Trends in Pharmacological Sciences*, 15: 437-441, 1994, and Tartaglia *et al.*, *Cell*, *83*:1263-1271, 1995. The isolation of the ChMIrp receptor(s) or ChMIrp binding partner(s) is useful for identifying or developing novel agonists and antagonists of the ChMIrp polypeptide-signaling pathway.

In another aspect of the present invention, the ChMIrp polypeptides may be used for identifying binding partners thereof ("ChMIrp binding partners" such as ChMIrp receptors and other ChMIrp cofactors). Yeast two-hybrid screens have been extensively used to identify and clone binding partners and receptors for proteins. (Chien *et al.*, *Proc. Natl. Acad. Sci. USA, 88*:9578-9583, 1991) The isolation of a ChMIrp binding partner(s) is useful for identifying or developing novel agonists and antagonists of the ChMIrp activity.

Such agonists and antagonists include soluble ChMIrp ligand(s), anti-h2520-109 selective binding agents (such as ChMIrp antibodies and derivatives thereof), small molecules, peptides or derivatives thereof capable of binding ChMIrp polypeptides, or antisense oligonucleotides, any of which can be used for

potentially treating one or more diseases or disorders, including those recited herein.

In certain embodiments, a ChMIrp polypeptide agonist or antagonist may be a protein, peptide, carbohydrate, lipid, or small molecular weight molecule which interacts with ChMIrp polypeptide to regulate its activity.

DESCRIPTION OF FIGURES

Figure 1 presents the polynucleotide sequence set out as SEQ ID NO: 3 which represents the entire coding region of the murine ChMIrp cDNA including the open reading frame (SEQ ID NO: 4).

Figure 2 presents the polynucleotide sequence set out as SEQ ID NO: 1 which represents the entire coding region of the human ChMIrp cDNA including the open reading frame (SEQ ID NO: 2).

Figure 3 shows the alignment of the human ChMIrp polypeptide sequence (SEQ ID NO: 2) and the mouse ChMIrp polypeptide sequence (SEQ ID NO: 4).

Figure 4 shows the homology between the human and mouse ChMIrp polypeptide sequences (SEQ ID NOS: 2 and 4; respectively) and the chondromodulin-I polypeptide sequence from various mammalian species (SEQ ID NOS: 5-9).

Figure 5 presents the complete genomic DNA sequence of human ChMIrp set out as SEQ ID NO: 10 with the exons underlined and the splice acceptor/donor sights in bold.

DETAILED DESCRIPTION OF THE INVENTION

The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described therein. All references cited in this application are expressly incorporated by reference herein.

Definitions:

The term "ChMIrp encoding nucleic acid molecule" refers to a nucleic acid molecule or polynucleotide comprising or consisting essentially of a nucleotide sequence set forth in SEQ ID NO: 1, and/or comprising or consisting essentially of a nucleotide sequence encoding the polypeptide as set forth in SEQ ID NO: 2. Related nucleic acid molecules comprise or consist essentially of a nucleotide sequence that is about 70 percent identical to the nucleotide sequence as shown in SEQ ID NO: 1, or comprise or consist essentially of a nucleotide sequence encoding a polypeptide that is about 75 percent identical to the polypeptide as set forth in SEQ ID NO: 2. In preferred embodiments, the nucleotide sequences are about 75 percent, or about 80 percent, or about 90 percent, or about 95, 96, 97, 98 or 99 percent identical to the nucleotide sequence as set forth in SEQ ID NO: 1, or the nucleotide sequences encoding a polypeptide that is about 75 percent, or about 80 percent, or about 85 percent, or about 90 percent, or about 95, 96, 97, 98, or 99 percent identical to the polypeptide sequence as set forth in SEQ ID NO: 2.

Related nucleic acid molecules also include fragments of the above ChMIrp nucleic acid molecules which are about 10 contiguous nucleotides, or about 15, or about 20, or about 25, or about 50, or about 75, or about 100 or greater than 100 contiguous nucleotides. Related nucleic acid molecules also include fragments of the above ChMIrp nucleic acid molecules which encode a polypeptide of at least about 25 amino acid residues, or about 50, or about 75, or about 100, or greater than about 100 amino acid residues. Related nucleic acid molecules also include a nucleotide sequence encoding a polypeptide comprising or consisting essentially of a substitution and/or deletion of one or more of amino acids 1-317 with reference to the polypeptide as set forth in SEQ ID NO: 2. Related ChMIrp nucleic acid molecules include those molecules which comprise nucleotide sequences which hybridize under moderate or highly stringent conditions as defined herein with any of the above nucleic acid molecules. In preferred embodiments, the related nucleic acid molecule comprise sequences which hybridize under moderate or highly stringent conditions with the sequence as shown in SEQ ID NO: 1, or with a molecule encoding a polypeptide, which polypeptide comprises the sequence as

shown in SEQ ID NO: 2, or with a nucleic acid fragment as defined above, or with a nucleic acid fragment encoding a polypeptide as defined above. It is also understood that related nucleic acid molecules include allelic or splice variants of any of the above nucleic acids, and include sequences which are complimentary to any of the above nucleotide sequences.

The term "isolated nucleic acid molecule" refers to a nucleic acid molecule of the invention that (1) has been separated from at least about 50 percent of proteins, lipids, carbohydrates or other materials with which it is naturally found when total DNA is isolated from the source cells, (2) is not linked to all or a poartion of a polynucleotide to which the "isolated nucleic acid molecule" is linked in nature, (3) is operably linked to a polynucleotide which it is not linked to in nature, or (4) does not occur in nature as part of a larger polynucleotide sequence. Preferably, the isolated nucleic acid of the present invention is substantially free from any other contaminating nucleic acid molecule(s) or other contaminants that are found in its natural environment that would interfere with its use in polypeptide production or its therapeutic, diagnostic, prophylactic or research use.

A "nucleic acid" sequence or molecule as used herein refers to a DNA or RNA sequence. The term encompasses molecules formed from any of the known base analogs of DNA and RNA such as, but not limited to 4-acetylcytosine, 8-hydroxy-N6-methyladenosine, aziridinylcytosine, pseudoisocytosine, 5-(carboxyhydroxylmethyl) uracil, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxy-methylamino-methyluracil, dihydrouracil, inosine, N6-iso-pentenyladenine, 1-methyladenine, 1-methylpseudouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethyl-guanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-methyl-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyamino-methyl-2-thiouracil, beta-D-mannosylqueosine, 5' -methoxycarbonyl-methyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, oxybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, N-

uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, pseudouracil, queosine, 2-thiocytosine, and 2,6-diaminopurine.

The term "operably linked" is used herein to refer to a method of flanking sequences wherein the flanking sequences so described are configured or assembled so as to perform their usual function. Thus, a flanking sequence operably linked to a coding sequence may be capable of effecting the replication, transcription and/or translation of the coding sequence. For example, a coding sequence is operably linked to a promoter when the promoter is capable of directing transcription of that coding sequence. A flanking sequence need not be contiguous with the coding sequence, so long as it functions correctly. Thus, for example, intervening untranslated yet transcribed sequences can be present between a promoter sequence and the coding sequence and the promoter sequence can still be considered "operably linked" to the coding sequence.

The term "ChMIrp polypeptide allelic variant" refers to one of several possible naturally occurring alternate forms of a gene occupying a given locus on a chromosome of an organism or population of organisms.

The term "ChMIrp polypeptide splice variant" refers to a nucleic acid molecule, usually RNA, which is generated by alternative processing of intron sequences in an RNA transcript.

The term "expression vector" refers to a vector which is suitable for propagation in a host cell and contains nucleic acid sequences which direct and/or control the expression of inserted heterologous nucleic acid sequences. Expression includes, but is not limited to, processes such as transcription, translation, and RNA splicing, if introns are present.

The term "vector" is used to refer to any molecule (*e.g.*, nucleic acid, plasmid, or virus) used to transfer coding information to a host cell.

The term "transformation" as used herein refers to a change in a cell's genetic characteristics, and a cell has been transformed when it has been modified to contain a new DNA. For example, a cell is transformed where it is genetically modified from its native state. Following transfection or transduction, the transforming DNA may recombine with that of the cell by physically integrating

into a chromosome of the cell, may be maintained transiently as an episomal element without being replicated, or may replicate independently as a plasmid. A cell is considered to have been stably transformed when the DNA is replicated with the division of the cell.

The term "transfection" is used to refer to the uptake of foreign or exogenous DNA by a cell, and a cell has been "transfected" when the exogenous DNA has been introduced inside the cell membrane. A number of transfection techniques are well known in the art and are disclosed herein. *See*, for example, Graham *et al.*, *Virology*, 52: 456, 1973; Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratories (Cold Spring Harbor, New York, 1989); Davis *et al.*, *Basic Methods in Molecular Biology*, Elsevier, 1986; and Chu *et al.*, *Gene*, 13: 197, 1981. Such techniques can be used to introduce one or more exogenous DNA moieties into suitable host cells.

The term "transduction" is used to refer to the transfer of genes from one bacterium to another, usually by a phage. "Transduction" also refers to the acquisition and transfer of eukaryotic cellular sequences by retroviruses.

The term "host cell" is used to refer to a cell which has been transformed, or is capable of being transformed, by a vector bearing a selected gene of interest which is then expressed by the cell. The term includes the progeny of the parent cell, whether or not the progeny is identical in morphology or in genetic make-up to the original parent, so long as the selected gene is present.

The term "vector" is used to refer to any molecule (*e.g.*, nucleic acid, plasmid, or virus) used to transfer coding information to a host cell.

The term "highly stringent conditions" refers to those conditions that are designed to permit hybridization of DNA strands whose sequences are highly complementary, and to exclude hybridization of significantly mismatched DNAs. Hybridization stringency is principally determined by temperature, ionic strength, and the concentration of denaturing agents such as formamide. Examples of "highly stringent conditions" for hybridization and washing are 0.015 M sodium chloride, 0.0015 M sodium citrate at 65-68°C or 0.015 M sodium chloride, 0.0015M sodium citrate, and 50% formamide at 42°C. See Sambrook *et al.*,

Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, (Cold Spring Harbor, N.Y. 1989); Anderson *et al.*, *Nucleic Acid Hybridisation: A Practical Approach*, Ch. 4, IRL Press Limited (Oxford, England).

More stringent conditions (such as higher temperature, lower ionic strength, higher formamide, or other denaturing agent) may also be used, however, the rate of hybridization will be affected. Other agents may be included in the hybridization and washing buffers for the purpose of reducing non-specific and/or background hybridization. Examples are 0.1% bovine serum albumin, 0.1% polyvinyl-pyrrolidone, 0.1% sodium pyrophosphate, 0.1% sodium dodecylsulfate, $NaDodSO_4$, (SDS), ficoll, Denhardt's solution, sonicated salmon sperm DNA (or other non-complementary DNA), and dextran sulfate, although other suitable agents can also be used. The concentration and types of these additives can be changed without substantially affecting the stringency of the hybridization conditions. Hybridization experiments are usually carried out at pH 6.8-7.4, however, at typical ionic strength conditions, the rate of hybridization is nearly independent of pH. *See* Anderson *et al.*, *Nucleic Acid Hybridisation: A Practical Approach*, Ch. 4, IRL Press Limited (Oxford, England).

Factors affecting the stability of DNA duplex include base composition, length, and degree of base pair mismatch. Hybridization conditions can be adjusted by one skilled in the art in order to accommodate these variables and allow DNAs of different sequence relatedness to form hybrids. The melting temperature of a perfectly matched DNA duplex can be estimated by the following equation:

$$T_m(°C) = 81.5 + 16.6(\log[Na+]) + 0.41(\%G+C) - 600/N - 0.72(\%\text{formamide})$$

where N is the length of the duplex formed, [Na+] is the molar concentration of the sodium ion in the hybridization or washing solution, %G+C is the percentage of (guanine+cytosine) bases in the hybrid. For imperfectly matched hybrids, the melting temperature is reduced by approximately 1°C for each 1% mismatch.

The term "moderately stringent conditions" refers to conditions under which a DNA duplex with a greater degree of base pair mismatching than could occur under "highly stringent conditions" is able to form. Examples of typical

"moderately stringent conditions" are 0.015 M sodium chloride, 0.0015 M sodium citrate at 50-65°C or 0.015 M sodium chloride, 0.0015 M sodium citrate, and 20% formamide at 37-50°C. By way of example, a "moderately stringent" condition of 50°C in 0.015 M sodium ion will allow about a 21% mismatch.

It will be appreciated by those skilled in the art that there is no absolute distinction between "highly" and "moderately" stringent conditions. For example, at 0.015 M sodium ion (no formamide), the melting temperature of perfectly matched long DNA is about 71°C. With a wash at 65°C (at the same ionic strength), this would allow for approximately a 6% mismatch. To capture more distantly related sequences, one skilled in the art can simply lower the temperature or raise the ionic strength.

A good estimate of the melting temperature in 1 M NaCl* for oligonucleotide probes up to about 20 nucleotides is given by:

Tm = 2°C per A-T base pair + 4°C per G-C base pair

*The sodium ion concentration in 6x salt sodium citrate (SSC) is 1 M. See Suggs *et al.*, *Developmental Biology Using Purified Genes*, p. 683, Brown and Fox (eds.) (1981).

High stringency washing conditions for oligonucleotides are usually at a temperature of 0-5°C below the Tm of the oligonucleotide in 6X SSC, 0.1% SDS.

The term "ChMIrp polypeptide" refers to a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, and related polypeptides described herein. Related polypeptides include: ChMIrp polypeptide allelic variants, ChMIrp polypeptide analogs, ChMIrp polypeptide splice variants, ChMIrp polypeptide variants, ChMIrp polypeptide fragments, ChMIrp polypeptides derivatives, substitutions, deletions, and/or insertion variants, ChMIrp fusion polypeptides, and ChMIrp polypeptide orthologs.

The term "ChMIrp polypeptide analog" refers to a related polypeptide with a similar amino acid sequence tot eh ChMIrp amino acid sequence set forth as SEQ ID NO: 2 with conserved and non-conserved amino acid substitutions.

The term "ChMIrp ortholog" refers to a polypeptide from another species that corresponds to ChMIrp polypeptide amino acid sequence set forth in SEQ ID NO: 2. For example, mouse and human ChMIrp polypeptides are considered orthologs of each other.

ChMIrp polypeptides may be mature polypeptides, as defined herein, and may or may not have an amino terminal methionine residue, depending on the method by which they are prepared. The term "ChMIrp polypeptide fragment" refers to a peptide or polypeptide that comprises less than the full length amino acid sequence of a ChMIrp polypeptide as set forth in SEQ ID NO: 2. Such a fragment may arise, for example, from a truncation at the amino terminus, a truncation at the carboxy terminus, and/or an internal deletion of the amino acid sequence. ChMIrp fragments may result from alternative RNA splicing or from *in vivo* protease activity.

The term "ChMIrp polypeptide fragment" refers to a polypeptide that comprises less than the full length amino acid sequence of a ChMIrp polypeptide as set forth in SEQ ID NO: 2. Such ChMIrp fragments can be 6 amino acids or more in length, and may arise, for example, from a truncation at the amino terminus (with or without a leader sequence), a truncation at the carboxy terminus, and/or an internal deletion of one or more residues from the amino acid sequence. ChMIrp fragments may result from alternative RNA splicing or from *in vivo* protease activity. Membrane-bound forms of a ChMIrp polypeptide are also contemplated by the present invention. In preferred embodiments, truncations and/or deletions comprise about 10 amino acids, or about 20 amino acids, or about 50 amino acids, or about 75 amino acids, or about 100 amino acids, or more than about 100 amino acids. The polypeptide fragments so produced will comprise about 25 contiguous amino acids, or about 50 amino acids, or about 75 amino acids, or about 100 amino acids, or about 150 amino acids, or about 200 amino acids. Such ChMIrp polypeptide fragments may optionally comprise an amino terminal methionine residue. It will be appreciated that such fragments can also be used, for example, to generate antibodies to ChMIrp polypeptides.

The term "ChMIrp polypeptide variants" refers to ChMIrp polypeptides comprising amino acid sequences which contain one or more amino acid sequence substitutions, deletions, and/or additions as compared to the ChMIrp polypeptide amino acid sequence set forth in SEQ ID NO: 2. Variants may be naturally occurring or artificially constructed using recombinant DNA technology. Such ChMIrp polypeptide variants may be prepared from the corresponding nucleic acid molecules encoding said variants, which have a DNA sequence that varies accordingly from the DNA sequences for wild type ChMIrp polypeptides as set forth in SEQ ID NO: 1.

One skilled in the art will be able to determine suitable variants of the native ChMIrp polypeptide using well known techniques. For example, one may predict suitable areas of the molecule that may be changed without destroying biological activity. Also, one skilled in the art will realize that even areas that may be important for biological activity or for structure may be subject to conservative amino acid substitutions without destroying the biological activity or without adversely affecting the polypeptide structure.

For predicting suitable areas of the molecule that may be changed without destroying activity, one skilled in the art may target areas not believed to be important for activity. For example, when similar polypeptides with similar activities from the same species or from other species are known, one skilled in the art may compare the amino acid sequence of ChMIrp polypeptide to such similar polypeptides. After making such a comparison, one skilled in the art can determine residues and portions of the molecules that are conserved among similar polypeptides. One skilled in the art would know that changes in areas of the ChMIrp molecule that are not conserved would be less likely to adversely affect the biological activity and/or structure of a ChMIrp polypeptide. One skilled in the art would also know that, even in relatively conserved regions, one may substitute chemically similar amino acids for the naturally occurring residues while retaining activity (conservative amino acid residue substitutions).

Additionally, one skilled in the art can review structure-function studies identifying residues in similar polypeptides that are important for activity or

structure. In view of such a comparison, one skilled in the art can predict the importance of amino acid residues in a ChMIrp polypeptide that correspond to amino acid residues that are important for activity or structure in similar polypeptides. One skilled in the art may opt for chemically similar amino acid substitutions for such predicted important amino acid residues of ChMIrp polypeptides.

If available, one skilled in the art can also analyze the three-dimensional structure and amino acid sequence in relation to that structure in similar polypeptides. In view of that information, one skilled in the art may predict the alignment of amino acid residues of ChMIrp polypeptide with respect to its three dimensional structure. One skilled in the art may choose not to make radical changes to amino acid residues predicted to be on the surface of the protein, since such residues may be involved in important interactions with other molecules.

ChMIrp polypeptide analogs of the invention can be determined by comparing the amino acid sequence of ChMIrp polypeptide with related family members. Exemplary ChMIrp polypeptide related family members include, but are not limited to, human chondromoldulin I, murine ChMIrp, murine chondromodulin I, rat chondromodulin I, bovine chondromodulin I, rabbit chondromodulin I. This comparison can be accomplished by using a Pileup alignment (Wisconsin GCG Program Package) or an equivalent (overlapping) comparison with multiple family members within conserved and non-conserved regions.

As shown in Figure 4, the predicted amino acid sequence of human ChMIrp polypeptide (SEQ ID NO: 2) is aligned with murine ChMIrp, murine chondromodulin I, rat chondromodulin I, bovine chondromodulin I, human chondromodulin I, rabbit chondromodulin I. (SEQ ID NOS: 4-9). Other ChMIrp polypeptide analogs can be determined using these or other methods known to those of skill in the art. These overlapping sequences provide guidance for conservative and non-conservative amino acids substitutions resulting in additional ChMIrp analogs. It will be appreciated that these amino acid substitutions can consist of naturally occurring or non-naturally occurring amino acids. For example, as depicted in Figure 4, alignment of the of these related polypeptides indicates

et al., *J. Mol. Biol.*, 157: 105-131, 1982). It is known that certain amino acids may be substituted for other amino acids having a similar hydropathic index or score and still retain a similar biological activity. In making changes based upon the hydropathic index, the substitution of amino acids whose hydropathic indices are within ±2 is preferred, those which are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity, particularly where the biologically functional equivalent protein or peptide thereby created is envisioned for use in immunological embodiments, as in the present case.

U.S. Patent No. 4,554,101 states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with its immunogenicity and antigenicity, *i.e.*, with a biological property of the protein. As detailed in U.S. Patent No. 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 ± 1); glutamate (+3.0 ± 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5 ± 1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4).

In making changes based upon similar hydrophilicity values, the substitution of amino acids whose hydrophilicity values are within ±2 is preferred, those which are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred. U.S. Patent No. 4,554,101 also teaches the identification and preparation of epitopes from primary amino acid sequences on the basis of hydrophilicity. Through the methods disclosed in U.S. Patent No. 4,554,101 one of skill in the art is able to identify epitopes from within a given amino acid sequence. These regions are also referred to as "epitopic core regions".

Numerous scientific publications have been devoted to the prediction of secondary structure, and to the identification of epitopes, from analyses of amino acid sequences. See Chou *et al.*, *Biochemistry*, 13(2): 222-245, 1974; Chou *et al.*,

Biochemistry, 113(2): 211-222, 1974; Chou *et al.*, *Adv. Enzymol. Relat. Areas Mol. Biol.*, 47: 45-148, 1978; Chou *et al.*, *Ann. Rev. Biochem.*, 47: 251-276, 1978 and Chou *et al.*, *Biophys. J.*, 26: 367-384, 1979. Moreover, computer programs are currently available to assist with predicting antigenic portions and epitopic core regions of proteins. Examples include those programs based upon the Jameson-Wolf analysis (Jameson *et al.*, *Comput. Appl. Biosci.*, 4(1): 181-186, 1998 and Wolf *et al.*, *Comput. Appl. Biosci.*, 4(1): 187-191, 1988); the program PepPlot (Brutlag *et al.*, *CABS*, 6: 237-245, 1990 and Weinberger *et al.*, *Science*, 228: 740-742, 1985) and other new programs for protein tertiary structure prediction (Fetrow *et al.*, *Biotechnology*, 11: 479-483, 1993).

Computer programs are also currently available to assist with predicting secondary structure. One method of predicting secondary structure is based upon homology modeling. For example, two polypeptides or proteins which have a sequence identity of greater than 30%, or similarity greater than 40% often have similar structural topologies. The recent growth of the protein structural data base (PDB) has provided enhanced predictability of secondary structure, including the potential number of folds within a polypeptide's or protein's structure. See Holm *et al.*, *Nucl. Acid. Res.*, *27*(1):244-247, 1999). It has been suggested (Brenner *et al.*, *Curr. Opin. Struct. Biol.*, *7*(3):369-376, 1997) that there are a limited number of folds in a given polypeptide or protein and that once a critical number of structures have been resolved, structural prediction will become dramatically more accurate.

Additional methods of predicting secondary structure include "threading" (Jones *et al.*, *Current Opin. Struct. Biol.*, *7(3)*:377-87, 1997; Sippl *et al.*, *Structure, 4(1)*:15-9, 1996), "profile analysis" (Bowie *et al.*, *Science, 253*:164-170, 1991; Gribskov *et al.*, *Meth. Enzym., 183*:146-159, 1990; Gribskov *et al.*, *Proc. Nat. Acad. Sci., 84(13)*:4355-4358, 1987), and "evolutionary linkage" (See Home, *supra*, and Brenner, *supra* 1997).

In preferred embodiments, the variants have from 1 to 3, or from 1 to 5, or from 1 to 10, or from 1 to 15, or from 1 to 20, or from 1 to 25, or from 1 to 50, or from 1 to 75, or from 1 to 100, or more than 100 amino acid substitutions,

insertions, additions and/or deletions, wherein the substitutions may be conservative, as described herein, or non-conservative, or any combination thereof. In addition, the variants can have additions of amino acid residues either at the carboxy terminus or at the amino terminus (with or without a leader sequence).

Preferred ChMIrp polypeptide variants include glycosylation variants wherein the number and/or type of glycosylation sites has been altered compared to native ChMIrp polypeptide. In one embodiment,ChMIrp polypeptide variants comprise a greater or a lesser number of N-linked glycosylation sites. An N-linked glycosylation site is characterized by the sequence: Asn-X-Ser or Thr, wherein the amino acid residue designated as X may be any amino acid residue except proline. The substitution(s) of amino acid residues to create this sequence provides a potential new site for the addition of an N-linked carbohydrate chain. Alternatively, substitutions which eliminate this sequence will remove an existing N-linked carbohydrate chain. Also provided is a rearrangement of N-linked carbohydrate chains wherein one or more N-linked glycosylation sites (typically those that are naturally occurring) are eliminated and one or more new N-linked sites are created. Additional preferred ChMIrp variants include cysteine variants, wherein one or more cysteine residues are deleted or substituted with another amino acid (*e.g.*, serine). Cysteine variants are useful when ChMIrp polypeptides must be refolded into a biologically active conformation such as after the isolation of insoluble inclusion bodies. Cysteine variants generally have fewer cysteine residues than the native protein, and typically have an even number to minimize interactions resulting from unpaired cysteines.

The term "ChMIrp fusion polypeptide" refers to a fusion of ChMIrp polypeptide, fragment, and/or variant thereof, with a heterologous peptide or polypeptide. Heterologous peptides and polypeptides include, but are not limited to: an epitope to allow for the detection and/or isolation of a ChMIrp fusion polypeptide; a transmembrane receptor protein or a portion thereof, such as an extracellular domain, or a transmembrane and intracellular domain; a ligand or a portion thereof which binds to a transmembrane receptor protein; an enzyme or portion thereof which is catalytically active; a polypeptide or peptide which

promotes oligomerization, such as a leucine zipper domain; a polypeptide or peptide which increases stability, such as an immunoglobulin constant region, and a polypeptide which has a therapeutic activity different from the amino acid sequence set forth as SEQ ID NO: 2 or a ChMIrp polypeptide variant.

In addition, a ChMIrp polypeptide may be fused to itself or to a fragment, variant, or derivative thereof. Fusions can be made either at the amino terminus or at the carboxy terminus of a ChMIrp polypeptide. Fusions may be direct with no linker or adapter molecule or may be through a linker or adapter molecule, such as one or more amino acid residues up to about 20 amino acids residues, or up to about 50 amino acid residues. A linker or adapter molecule may also be designed with a cleavage site for a DNA restriction endonuclease or for a protease to allow for the separation of the fused moieties. It will be appreciated that once constructed, the fusion polypeptides can be derivatized according to the methods described herein.

In a further embodiment of the invention, a ChMIrp polypeptide, including a fragment, variant, and/or derivative, is fused to an Fc region of human IgG. Antibodies comprise two functionally independent parts, a variable domain known as "Fab", which binds antigen, and a constant domain known as "Fc", which links to such effector functions as complement activation and attack by phagocytic cells. An Fc has a long serum half-life, whereas an Fab is short-lived (Capon *et al.*, *Nature,* 337: 525-31, 1989). When constructed together with a therapeutic protein a Fc domain can provide longer half-life or incorporate such functions as Fc receptor binding, protein binding, complement fixation and perhaps even placental transfer. *Id.* Table I summarizes the use of certain Fc fusions known in the art, including materials and methods applicable to the production of fused ChMIrp polypeptides.

Table I
Fc fusion with Therapeutic Proteins

Form of Fc	Fusion partner	Therapeutic implications	Reference
IgG1	N-terminus of CD30-L	Hodgkin's disease; anaplastic lymphoma; T-cell leukemia	U.S. Patent No. 5,480,981
Murine Fcg2a	IL-10	anti-inflammatory; transplant rejection	Zheng *et al.*, *J. Immunol.*, 154: 5590-600, 1995
IgG1	TNF receptor	septic shock	Fisher *et al.*, *N. Engl. J. Med.*, 334: 1697-1702, 1996; Van Zee *et al.*, *J. Immunol.*, 156: 2221-30, 1996
IgG, IgA, IgM, or IgE (excluding the first domain)	TNF receptor	inflammation, autoimmune disorders	U.S. Pat. No. 5,808,029, issued September 15, 1998
IgG1	CD4 receptor	AIDS	Capon *et al.*, *Nature* 337: 525-31, 1996
IgG1, IgG3	N-terminus of IL-2	anti-cancer, antiviral	Harvill *et al.*, *Immunotech.*, 1: 95-105, 1996
IgG1	C-terminus of OPG	osteoarthritis; bone density	WO 97/23614, published July 3, 1997
IgG1	N-terminus of leptin	anti-obesity	PCT/US 97/23183, filed December 11, 1997
Human Ig Cg1	CTLA-4	autoimmune disorders	Linsley (1991), *J. Exp. Med.*, 174:561-9

In one example, a human IgG hinge, CH2 and CH3 region may be fused at either the N-terminus or C-terminus of the ChMIrp polypeptides using methods known to the skilled artisan. In another example, a portion of a hinge regions and CH2 and CH3 regions may be fused. The resulting ChMIrp Fc-fusion polypeptide may be purified by use of a Protein A affinity column. Peptides and proteins fused to a Fc region have been found to exhibit a substantially greater half-life *in vivo* than the unfused counterpart. Also, a fusion to an Fc region allows for

dimerization/multimerization of the fusion polypeptide. The Fc region may be a naturally occurring Fc region, or may be altered to improve certain qualities, such as therapeutic qualities, circulation time, reduce aggregation, etc.

The term "ChMIrp polypeptide derivatives" refers to ChMIrp polypeptides, variants, or fragments thereof, that have been chemically modified, as for example, by covalent attachment of one or more water soluble polymers, N-linked or O-linked carbohydrates, sugars, phosphates, and/or other such molecules. Such modifications may be introduced into the molecule by reacting targeted amino acid residues of the purified or crude protein with an organic derivatizing agent that is capable of reacting with selected side chains or terminal residues. The resulting covalent derivatives are also useful in programs directed at identifying residues important for biological activity. The derivatives are modified in a manner that is different from naturally occurring ChMIrp polypeptide either in the type or location of the molecules attached to the polypeptide. Derivatives further include deletion of one or more chemical groups naturally attached to the ChMIrp polypeptide.

For example, the polypeptides may be modified by the covalent attachment of one or more polymers, including, but not limited to, water soluble polymers, N-linked or O-linked carbohydrates, sugars, phosphates, and/or other such molecules. For example, the polymer selected is typically water soluble so that the protein to which it is attached does not precipitate in an aqueous environment, such as a physiological environment. The polymer may be of any molecular weight, and may be branched or unbranched. Included within the scope of suitable polymers is a mixture of polymers. Preferably, for therapeutic use of the end-product preparation, the polymer will be pharmaceutically acceptable.

Suitable water soluble polymers or mixtures thereof include, but are not limited to, polyethylene glycol (PEG), monomethoxy-polyethylene glycol, dextran (such as low molecular weight dextran, of, for example about 6 kD), cellulose, or other carbohydrate based polymers, poly-(N-vinyl pyrrolidone) polyethylene glycol, propylene glycol homopolymers, a polypropylene oxide/ethylene oxide co-polymer, polyoxyethylated polyols (*e.g.*, glycerol) and

polyvinyl alcohol. Also encompassed by the present invention are bifunctional PEG crosslinking molecules which may be used to prepare covalently attached ChMIrp multimers.

For the acylation reactions, the polymer(s) selected should have a single reactive ester group. For reductive alkylation, the polymer(s) selected should have a single reactive aldehyde group. A reactive aldehyde is, for example, polyethylene glycol propionaldehyde, which is water stable, or mono C_1-C_{10} alkoxy or aryloxy derivatives thereof (*see* U.S. Patent No. 5,252,714).

The pegylation of ChMIrp polypeptides may be carried out by any of the pegylation reactions known in the art, as described for example in the following references: Francis *et al.*, *Focus on Growth Factors*, 3: 4-10, 1992; EP 0154316; EP 0401384 and U.S. Patent No. 4,179,337. Pegylation may be carried out via an acylation reaction or an alkylation reaction with a reactive polyethylene glycol molecule (or an analogous reactive water-soluble polymer) as described herein. Polyethylene glycol (PEG) is a water-soluble polymer suitable for use herein. As used herein, the terms "polyethylene glycol" and "PEG" are meant to encompass any of the forms of PEG that have been used to derivatize proteins, including mono-(C_1-C_{10}) alkoxy- or aryloxy-polyethylene glycol.

In general, chemical derivatization may be performed under any suitable conditions used to react a biologically active substance with an activated polymer molecule. Methods for preparing pegylated ChMIrp polypeptides will generally comprise the steps of (a) reacting the polypeptide with polyethylene glycol (such as a reactive ester or aldehyde derivative of PEG) under conditions whereby ChMIrp polypeptide becomes attached to one or more PEG groups, and (b) obtaining the reaction product(s). In general, the optimal reaction conditions for the acylation reactions will be determined based on known parameters and the desired result. For example, the larger the ratio of PEG:protein, the greater the percentage of poly-pegylated product. In one embodiment, the ChMIrp polypeptide derivative may have a single PEG moiety at the amino terminus. See, for example, U.S. Patent No. 5,234,784.

Generally, conditions which may be alleviated or modulated by the administration of the present ChMIrp polypeptide derivative include those described herein. However, the ChMIrp polypeptide derivative disclosed herein may have additional activities, enhanced or reduced biological activity, or other characteristics, such as increased or decreased half-life, as compared to the non-derivatized molecules.

The terms "biologically active ChMIrp polypeptides", "biologically active ChMIrp polypeptide fragments", "biologically active ChMIrp polypeptide variants", and "biologically active ChMIrp polypeptide derivatives" refer to ChMIrp polypeptides having at least one activity characteristic of a ChMIrp polypeptide. Immunogenic fragments of ChMIrp polypeptides are those capable of inducing in a host animal antibodies directed to the ChMIrp fragment.

"Naturally occurring" or "native" when used in connection with biological materials such as nucleic acid molecules, polypeptides, host cells, and the like, refers to materials which are found in nature and are not manipulated by man. Similarly, "non-naturally occurring" or "non-native" as used herein refers to a material that is not found in nature or that has been structurally modified or synthesized by man.

The term "isolated polypeptide" refers to a polypeptide of the present invention that (1) has been separated from at least about 50 percent of polynucleotides, lipids, carbohydrates or other materials with which it is naturally found when isolated from the source cell, (2) is not linked (by covalent or noncovalent interaction) to all or a portion of a polypeptide to which the "isolated polypeptide" is linked in nature, (3) is operably linked (by covalent or noncovalent interaction) to a polypeptide with which it is not linked in nature, or (4) does not occur in nature. Preferably, the isolated polypeptide is substantially free from any other contaminating polypeptides or other contaminants that are found in its natural environment that would interfere with its therapeutic, diagnostic, prophylactic or research use.

The term "mature ChMIrp polypeptide" refers to a polypeptide lacking a leader sequence and may also include other modifications of a polypeptide

such as proteolytic processing of the amino terminus (with or without a leader sequence) and/or the carboxy terminus, cleavage of a smaller polypeptide from a larger precursor, N-linked and/or O-linked glycosylation, and the like.

The term "mutein" refers to a mutant protein, polypeptide, variants, analogs or fragments of ChMIrp polypeptide. Muteins of ChMIrp may be prepared by deletion, insertion, substitution, point mutation, truncation, addition, transposition, PCR amplification, site-directed mutagenesis or other methods known in the art.

The terms "effective amount" and "therapeutically effective amount" refer to the amount of a ChMIrp polypeptide (or ChMIrp antagonist) necessary to support an observable change in the level of one or more biological activities of the ChMIrp polypeptides as set forth above, to bring about a meaningful patient benefit, *i.e.* treatment, healing, prevention, or amelioration of a condition. When applied to an individual active ingredient, administered alone, the term refers to that ingredient alone. When applied to combination, the term refers to combined amounts of active ingredients that result in therapeutic effect, when administered in combination, serially or simultaneously. The ChMIrp polypeptides that have use in practicing the present invention may be naturally occurring full length polypeptides, or truncated polypeptides or variant homologs or analogs or derivatives or peptide fragments. Illustrative analogs include those in which one or more divergent amino acids between two species are substituted with the divergent amino acid from another species. Divergent amino acids may also be substituted with any other amino acid whether it be a conservative or a non-conservative amino acid.

The terms "pharmaceutically acceptable carrier" or "physiologically acceptable carrier" as used herein refer to one or more formulation materials suitable for accomplishing or enhancing the delivery of the ChMIrp polypeptide,ChMIrp nucleic acids molecule, or ChMIrp selective binding agent as a pharmaceutical composition.

The term "selective binding agent" refers to a molecule or molecules having specificity for ChMIrp molecules. Selective binding agents include antibodies, such as polyclonal antibodies, monoclonal antibodies (mAbs), chimeric

antibodies, CDR-grafted antibodies, anti-idiotypic (anti-Id) antibodies to antibodies that can be labeled in soluble or bound form, as well as fragments, regions, or derivatives thereof which are provided by known techniques, including, but not limited to enzymatic cleavage, peptide synthesis, or recombinant techniques. The anti-ChMIrp selective binding agents of the present invention are capable, for example, of binding portions of ChMIrp molecules that inhibit the binding of ChMIrp molecules to ChMIrp receptors.

As used herein, the terms, "specific" and "specificity" refer to the ability of the selective binding agents to bind to human ChMIrp polypeptides. It will be appreciated, however, that the selective binding agents may also bind orthologs of ChMIrp polypeptides, that is, interspecies versions of ChMIrp polypeptides, such as mouse and rat ChMIrp polypeptides. A preferred embodiment relates to antibodies that are highly specific to ChMIrp polypeptides yet do not cross-react (that is, they fail to bind) with specificity to non-ChMIrp polypeptides.

The term "antigen" refers to a molecule or a portion of a molecule capable of being bound by a selective binding agent, such as an antibody, which is additionally capable of inducing an animal to produce antibodies capable of binding to an epitope of that antigen. An antigen can have one or more epitopes. The specific binding reaction referred to above is meant to indicate that the antigen will react, in a highly selective manner, with its corresponding antibody and not with the multitude of other antibodies which can be evoked by other antigens.

ChMIrp polypeptides, fragments, variants, and derivatives may be used to prepare ChMIrp selective binding agents using methods known in the art. Thus, antibodies and antibody fragments that bind ChMIrp polypeptides are within the scope of the present invention. Antibody fragments include those portions of the antibody which bind to an epitope on the ChMIrp polypeptide. Examples of such fragments include Fab and F(ab') fragments generated by enzymatic cleavage of full-length antibodies. Other binding fragments include those generated by recombinant DNA techniques, such as the expression of recombinant plasmids containing nucleic acid sequences encoding antibody variable regions. These

antibodies may be, for example, polyclonal monospecific polyclonal, monoclonal, recombinant, chimeric, humanized, human, single chain, and/or bispecific.

Relatedness of Nucleic Acid Molecules and/or Polypeptides

The term "identity", as known in the art, refers to a relationship between the sequences of two or more polypeptide molecules or two or more nucleic acid molecules, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or nucleic acid molecule sequences, as the case may be, as determined by the match between strings of nucleotide or amino acid sequences. "Identity" measures the percent of identical matches between two or more sequences with gap alignments addressed by particular computer programs (*i.e.*, "algorithms").

The term "similarity" is a related concept, but in contrast to "identity", refers to a measure of similarity which includes both identical matches and conservative substitution matches. Identity and similarity of related nucleic acid molecules and polypeptides can be readily calculated by known methods, including but not limited to those described in: *Computational Molecular Biology*, Lesk, A.M., *ed.*, Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., *ed.*, Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part 1, Griffin, A.M., and Griffin, H.G., *eds.*, Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer,* Gribskov, M. and Devereux, J., *eds.*, M. Stockton Press, New York, 1991; and Carillo and Lipman, *SIAM J. Applied Math.*, 48: 1073, 1988. Since conservative substitutions apply to polypeptides and not nucleic acid molecules, similarity only deals with polypeptide sequence comparisons. If two polypeptide sequences have, for example, 10/20 identical amino acids, and the remainder are all non-conservative substitutions, then the percent identity and similarity would both be 50%. If in the same example, there are 5 more positions where there are conservative substitutions, then the percent identity remains 50%, but the percent similarity would be 75% (15/20). Therefore, in cases where there are conservative

substitutions, the degree of similarity between two polypeptide sequences will be higher than the percent identity between those two sequences.

Differences in the nucleic acid sequence may result in conservative and/or non-conservative modifications of the amino acid sequence relative to the amino acid sequence of SEQ ID NO: 2.

The term "conservative amino acid substitution" refers to a substitution of a native amino acid residue with a nonnative residue such that there is little or no effect on the polarity or charge of the amino acid residue at that position. For example, a conservative substitution results from the replacement of a non-polar residue in a polypeptide with any other non-polar residue. Further, any native residue in the polypeptide may also be substituted with alanine, as has been previously described for "alanine scanning mutagenesis". General rules for amino acid substitutions are set forth in Table II below.

Table II

Amino Acid Substitutions

Original Residues	Exemplary Substitutions	Preferred Substitutions
Ala	Val, Leu, Ile	Val
Arg	Lys, Gln, Asn	Lys
Asn	Gln	Gln
Asp	Glu	Glu
Cys	Ser, Ala	Ser
Gln	Asn	Asn
Glu	Asp	Asn
Gly	Pro, Ala	Ala
His	Asn, Gln, Lys, Arg	Arg
Ile	Leu, Val, Met, Ala, Phe,	Leu
Leu	Norleucine, Ile, Val,	Leu
Lys	Arg, 1,4 Diaminobutyric	Arg
Met	Leu, Phe, Ile	Leu
Phe	Leu, Val, Ile, Ala, Tyr	Arg
Pro	Ala	Gly
Ser	Thr, Ala, Cys	Thr
Thr	Ser	Ser
Trp	Tyr, Phe	Tyr
Tyr	Trp, Phe, Thr, Ser	Phe
Val	Ile, Met, Leu, Phe, Ala,	Leu

Conservative modifications to the amino acid sequence (and the corresponding modifications to the encoding nucleotides) are expected to produce ChMIrp having functional and chemical characteristics similar to those of naturally occurring ChMIrp. In contrast, substantial modifications in the functional and/or chemical characteristics of ChMIrp may be accomplished by selecting substitutions that differ significantly in their effect on maintaining (a) the structure of the molecular backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or

(c) the bulk of the side chain. Naturally occurring residues may be divided into classes based on common side chain properties:

1) hydrophobic: norleucine, Met, Ala, Val, Leu, Ile;
2) neutral hydrophilic: Cys, Ser, Thr, Asn, Gln;
3) acidic: Asp, Glu;
4) basic: His, Lys, Arg;
5) residues that influence chain orientation: Gly, Pro; and
6) aromatic: Trp, Tyr, Phe.

Non-conservative substitutions may involve the exchange of a member of one of these classes for a member from another class. Such substituted residues may be introduced into regions of the human ChMIrp molecule that are homologous with non-human ChMIrp or into the non-homologous regions of the molecule.

Conservative amino acid substitutions also encompass non-naturally occurring amino acid residues which are typically incorporated by chemical peptide synthesis rather than by synthesis in biological systems. These include peptidomimetics and other reversed or inverted forms of amino acid moieties.

Preferred methods to determine identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package, including GAP (Devereux *et al.*, *Nucleic Acids Research* 12(1): 387, 1984); Genetics Computer Group, University of Wisconsin, Madison, WI), BLASTP, BLASTN, and FASTA (Atschul *et al.*, *J. Molec. Biol.* 215: 403-410, 1990). The BLAST X program is publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul, S., *et al.*, NCB NLM NIH Bethesda, MD 20894; Altschul *et al.*, *J. Mol. Biol.* 215: 403-410, 1990). The well known Smith Waterman algorithm may also be used to determine identity.

Certain alignment schemes for aligning two amino acid sequences may result in the matching of only a short region of the two sequences, and this small aligned region may have very high sequence identity even though there is no significant relationship between the two full length sequences. Accordingly, in a preferred embodiment, the selected alignment method (GAP program) will result in an alignment that spans at least 50 contiguous amino acids of the claimed polypeptide.

By way of example, using the computer algorithm GAP (Genetics Computer Group, University of Wisconsin, Madison, WI), two polypeptides for which the percent sequence identity is to be determined are aligned for optimal matching of their respective amino acids (the "matched span", as determined by the algorithm). A gap opening penalty (which is calculated as 3x the average diagonal; the "average diagonal" is the average of the diagonal of the comparison matrix being used; the "diagonal" is the score or number assigned to each perfect amino acid match by the particular comparison matrix) and a gap extension penalty (which is usually 1/10 times the gap opening penalty), as well as a comparison matrix such as PAM 250 or BLOSUM 62 are used in conjunction with the algorithm. A standard comparison matrix (Dayhoff *et al.*, In: *Atlas of Protein Sequence and Structure*, vol. 5, supp.3, 1978, for the PAM250 comparison matrix; *see* Henikoff *et al., Proc. Natl. Acad. Sci USA*, 89: 10915-10919, 1992, for the BLOSUM 62 comparison matrix) is also used by the algorithm.

Preferred parameters for polypeptide sequence comparison include the following:

Algorithm: Needleman and Wunsch, *J. Mol. Biol.* 48: 443-453, 1970,

Comparison matrix: BLOSUM 62 from Henikoff and Henikoff, *Proc. Natl. Acad. Sci.USA* 89: 10915-10919, 1992;

Gap Penalty: 12

Gap Length Penalty: 4

Threshold of Similarity: 0

The GAP program is useful with the above parameters. The aforementioned parameters are the default parameters for polypeptide comparisons (along with no penalty for end gaps) using the GAP algorithm.

Preferred parameters for nucleic acid molecule sequence comparison include the following:

Algorithm: Needleman and Wunsch, *J. Mol. Biol.* 48: 443-453, 1970;

Comparison matrix: matches = +10, mismatch = 0

Gap Penalty: 50

Gap Length Penalty: 3

The GAP program is also useful with the above parameters. The aforementioned parameters are the default parameters for nucleic acid molecule comparisons.

Other exemplary algorithms, gap opening penalties, gap extension penalties, comparison matrices, thresholds of similarity, etc. may be used by those of skill in the art, including those set forth in the Program Manual, Wisconsin Package, Version 9, September, 1997. The particular choices to be made will depend on the specific comparison to be made, such as DNA to DNA, protein to protein, protein to DNA; and additionally, whether the comparison is between pairs of sequences (in which case GAP or BestFit are generally preferred) or between one sequence and a large database of sequences (in which case FASTA or BLASTA are preferred).

Synthesis

It will be appreciated by those skilled in the art the nucleic acid and polypeptide molecules described herein may be produced by recombinant and other means.

Nucleic Acid Molecules

The nucleic acid molecules encode a polypeptide comprising the amino acid sequence of ChMIrp polypeptide and can readily be obtained in a variety

of ways including without limitation, chemical synthesis, cDNA or genomic library screening, expression library screening and/or PCR amplification of cDNA.

Recombinant DNA methods used herein are generally, but not limited to, those set forth in Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989) and/or Ausubel *et al., eds.*, (*Current Protocols in Molecular Biology*, Green Publishers Inc. and Wiley and Sons, NY, 1994).

The present invention provides for nucleic acid molecules as described herein and methods for obtaining the molecules. A gene or cDNA encoding a "ChMIrp polypeptide" or fragment thereof may be obtained by hybridization screening of a genomic or cDNA library, or by PCR amplification. Probes or primers useful for screening a library by hybridization can be generated based on sequence information for other known genes or gene fragments from the same or a related family of genes, such as, for example, conserved motifs.

Where a gene encoding ChMIrp polypeptide has been identified from one species, all or a portion of that gene may be used as a probe to identify corresponding genes from other species (orthologs) or related genes from the same species (homologs). The probes or primers may be used to screen cDNA libraries from various tissue sources believed to express the ChMIrp gene.

In addition, part or all of a nucleic acid molecule having the sequence as set forth in SEQ ID NO: 1 may be used to screen a genomic library to identify and isolate a gene encoding ChMIrp. Typically, conditions of moderate or high stringency will be employed for screening to minimize the number of false positives obtained from the screen. The availability of the cDNA coding for the ChMIrp or fractions thereof is the prerequisite for obtaining the genomic DNA. Under stringent conditions, a DNA library is screened and the clones obtained are investigated to see whether they contain the regulatory sequence elements needed for gene expression in addition to the coding regions (*e.g.* checking for promoter function by fusion with coding regions of suitable reporter genes). Methods for screening DNA libraries under stringent conditions are taught, for example, in published European Patent Application No. EPA 0 174 143. Obtaining the genomic DNA sequence makes it

possible to investigate the regulatory sequences situated in the area which does not code for the ChMIrp, particularly in the 5'-flanking region, for any possible interaction with known substances which modulate gene expression, *e.g.* transcription factors or steroids, or possibly discover new substances which might have a specific effect on the expression of this gene. The results of such investigations provide the basis for the targeted use of such substances for modulating ChMIrp expression and hence for directly influencing the ability of the cells to interact with chondromodulin receptors. As a result, the specific reaction with the receptors and the resulting effects can be suppressed.

The scope of the present invention also includes DNAs which code for subtypes of ChMIrp, which may have properties different from those of the present ChMIrp. These are expression products which are formed by alternative splicing and have modified structures in certain areas, *e.g.* structures which can bring about a change in the affinity and specificity for the receptor or a change in terms of the nature and efficiency of signal transmission.

With the aid of the cDNA coding for the ChMIrp it is possible to obtain nucleic acids which hybridize with the cDNA or fragments thereof under conditions of low stringency, moderate stringency or high stingency, and code for a polypeptide capable of binding receptors or contain the sequence coding for such a polypeptide.

Nucleic acid molecules encoding ChMIrp polypeptides may also be identified by expression cloning which employs detection of positive clones based upon a property of the expressed protein. Typically, nucleic acid libraries are screened by binding of an antibody or other binding partner (*e.g.*, receptor or ligand) to cloned proteins which are expressed and displayed on the host cell surface. The antibody or binding partner is modified with a detectable label to identify those cells expressing the desired clone.

Recombinant expression techniques conducted in accordance with the descriptions set forth below may be followed to produce these polynucleotides and to express the encoded polypeptides. For example, by inserting a nucleic acid sequence which encodes the amino acid sequence of a ChMIrp polypeptide into an appropriate

vector, one skilled in the art can readily produce large quantities of the desired nucleotide sequence. The sequences can then be used to generate detection probes or amplification primers. Alternatively, a polynucleotide encoding the amino acid sequence of a ChMIrp polypeptide can be inserted into an expression vector. By introducing the expression vector into an appropriate host, the encoded ChMIrp polypeptide may be produced in large amounts.

Another method for obtaining a suitable nucleic acid sequence is the polymerase chain reaction (PCR). In this method, cDNA is prepared from poly(A)+RNA or total RNA using the enzyme reverse transcriptase. Two primers, typically complementary to two separate regions of cDNA (oligonucleotides) encoding the amino acid sequence of a ChMIrp polypeptide, are then added to the cDNA along with a polymerase such as Taq polymerase, and the polymerase amplifies the cDNA region between the two primers.

Another means of preparing a nucleic acid molecule encoding the amino acid sequence of a ChMIrp polypeptide is by chemical synthesis using methods well known to the skilled artisan such as those described by Engels *et al.*, (Angew, *Chem. Intl. Ed.*, 28: 716-734, 1989). These methods include, *inter alia*, the phosphotriester, phosphoramidite, and H-phosphonate methods for nucleic acid synthesis. A preferred method for such chemical synthesis is polymer-supported synthesis using standard phosphoramidite chemistry. Typically, the DNA encoding the amino acid sequence of a ChMIrp polypeptide will be several hundred nucleotides in length. Nucleic acids larger than about 100 nucleotides can be synthesized as several fragments using these methods. The fragments can then be ligated together to form the full-length nucleotide sequence of a ChMIrp polynucleotide. Usually, the DNA fragment encoding the amino terminus of the polypeptide will have an ATG, which encodes a methionine residue. This methionine may or may not be present on the mature form of the ChMIrp polypeptide, depending on whether the polypeptide produced in the host cell is designed to be secreted from that cell.

In some cases, it may be desirable to prepare nucleic acid molecules encoding the ChMIrp polypeptide variants or muteins. Nucleic acid molecules

encoding variants may be produced using site directed mutagenesis, transposition, deletion, addition, truncation, PCR amplification, or other appropriate methods, where the primer(s) have the desired point mutations (*see* Sambrook *et al.*, *supra*, and Ausubel *et al.*, *supra*, for descriptions of mutagenesis techniques), provided that DNA's modified in this way code for polypeptides capable of binding one or more members of the chondromodulin-family. Chemical synthesis using methods described by Engels *et al.*, *supra*, may also be used to prepare such variants. Other methods known to the skilled artisan may be used as well.

In certain embodiments, nucleic acid variants contain codons which have been altered for the optimal expression of a ChMIrp polypeptide in a given host cell. Particular codon alterations will depend upon the ChMIrp polypeptide(s) and host cell(s) selected for expression. Such "codon optimization" can be carried out by a variety of methods, for example, by selecting codons which are preferred for use in highly expressed genes in a given host cell. Computer algorithms which incorporate codon frequency tables such as "Ecohigh.cod" for codon preference of highly expressed bacterial genes may be used and are provided by the University of Wisconsin Package Version 9.0, Genetics Computer Group, Madison, WI. Other useful codon frequency tables include "Celegans_high.cod", "Celegans_low.cod", "Drosophila_high.cod", "Human_high.cod", "Maize_high.cod", and "Yeast_high.cod".

In other embodiments, nucleic acid molecules encode ChMIrp variants with conservative amino acid substitutions as defined above, ChMIrp variants comprising an addition and/or a deletion of one or more N-linked or O-linked glycosylation sites, or ChMIrp polypeptide fragments as described above. In addition, nucleic acid molecules may encode any combination of ChMIrp variants, fragments, and fusion polypeptides described herein provided that DNA's modified in this way code for polypeptides capable of finding one or more members of chondromodulin family of ligands and receptors.

Vectors and Host Cells

A nucleic acid molecule encoding the amino acid sequences of ChMIrp polypeptide may be inserted into an appropriate expression vector using standard ligation techniques. The vector is typically selected to be functional in the particular host cell employed (*i.e.*, the vector is compatible with the host cell machinery such that amplification of the gene and/or expression of the gene can occur). A nucleic acid molecule encoding the amino acid sequence of ChMIrp polypeptide may be amplified and/or expressed in prokaryotic, yeast, insect (baculovirus systems) and/or eukaryotic host cells. Selection of the host cell will depend in part on whether the ChMIrp polypeptide is to be post-translationally modified (*e.g.*, glycosylated and/or phosphorylated). If so, yeast, insect, or mammalian host cells are preferable. For a review of expression vectors, see *Meth. Enz.* v.185, D.V. Goeddel, ed. Academic Press Inc., San Diego CA, (1990).

Typically, expression vectors used in any of the host cells will contain sequences for plasmid maintenance and for cloning and expression of exogenous nucleotide sequences. Such sequences, collectively referred to as "flanking sequences" in certain embodiments, will typically include one or more of the following nucleotide sequences: a promoter, one or more enhancer sequences, an origin of replication, a transcriptional termination sequence, a complete intron sequence containing a donor and acceptor splice site, a sequence encoding a leader sequence for polypeptide secretion, a ribosome binding site, a polyadenylation sequence, a polylinker region for inserting the nucleic acid encoding the polypeptide to be expressed, and a selectable marker element. Each of these sequences is discussed below.

Optionally, the vector may contain a "tag"-encoding sequence, *i.e.*, an oligonucleotide molecule located at the 5′ or 3′ end of the ChMIrp polypeptide coding sequence; the oligonucleotide sequence encodes polyHis (such as hexaHis), or another "tag" such as FLAG, HA (hemaglutinin influenza virus) or myc for which commercially available antibodies exist. This tag is typically fused to the polypeptide upon expression of the polypeptide, and can serve as a means for affinity purification of the ChMIrp polypeptide from the host cell. Affinity purification can be

accomplished, for example, by column chromatography using antibodies against the tag as an affinity matrix or metal affinity chromatography (such as nickel columns with an affinity to 6 his tags). Optionally, the tag can subsequently be removed from the purified ChMIrp polypeptide by various means such as using certain peptidases for cleavage.

Flanking sequences may be homologous (*i.e.*, from the same species and/or strain as the host cell), heterologous (*i.e.*, from a species other than the host cell species or strain), hybrid (*i.e.*, a combination of flanking sequences from more than one source), or synthetic, or the flanking sequences may be native sequences which normally function to regulate ChMIrp polypeptide expression. As such, the source of a flanking sequence may be any prokaryotic or eukaryotic organism, any vertebrate or invertebrate organism, or any plant, provided that the flanking sequences is functional in, and can be activated by, the host cell machinery.

The flanking sequences useful in the vectors of this invention may be obtained by any of several methods well known in the art. Typically, flanking sequences useful herein other than the endogenous ChMIrp gene flanking sequences will have been previously identified by mapping and/or by restriction endonuclease digestion and can thus be isolated from the proper tissue source using the appropriate restriction endonucleases. In some cases, the full nucleotide sequence of one or more flanking sequence may be known. Here, the flanking sequence may be synthesized using the methods described herein for nucleic acid synthesis or cloning.

Where all or only a portion of the flanking sequence is known, it may be obtained using PCR and/or by screening a genomic library with suitable oligonucleotide and/or flanking sequence fragments from the same or another species.

Where the flanking sequence is not known, a fragment of DNA containing a flanking sequence may be isolated from a larger piece of DNA that may contain, for example, a coding sequence or even another gene or genes. Isolation may be accomplished by restriction endonuclease digestion to produce the proper DNA fragment followed by isolation using agarose gel purification, Qiagen® column chromatography (Chatsworth, CA), or other methods known to the skilled artisan.

Other selection genes may be used to amplify the gene which will be expressed. Amplification is the process wherein genes which are in greater demand for the production of a protein critical for growth are reiterated in tandem within the chromosomes of successive generations of recombinant cells. Examples of suitable, selectable markers for mammalian cells include dihydrofolate reductase (DHFR) and thymidine kinase. The mammalian cell transformants are placed under selection pressure which only the transformants are uniquely adapted to survive by virtue of the selection gene present in the vector. Selection pressure is imposed by culturing the transformed cells under conditions in which the concentration of selection agent in the medium is successively changed, thereby leading to the amplification of both the selection gene and the DNA that encodes ChMIrp polypeptide. As a result, increased quantities of ChMIrp polypeptide are synthesized from the amplified DNA.

A ribosome binding site is usually necessary for translation initiation of mRNA and is characterized by a Shine-Dalgarno sequence (prokaryotes) or a Kozak sequence (eukaryotes). The element is typically located 3′ to the promoter and 5′ to the coding sequence of the ChMIrp polypeptide to be expressed. The Shine-Dalgarno sequence is varied but is typically a polypurine (*i.e.*, having a high A-G content). Many Shine-Dalgarno sequences have been identified, each of which can be readily synthesized using methods set forth herein and used in a prokaryotic vector.

A leader, or signal, sequence may be used to direct the secretion of ChMIrp polypeptide out of the host cell where it is synthesized. Typically, the signal sequence is positioned in the coding region of the ChMIrp nucleic acid molecule, or directly at the 5′ end of the ChMIrp polypeptide coding region. Many signal sequences have been identified, and any of them that are functional in the selected host cell may be used in conjunction with the ChMIrp gene or cDNA. Therefore, a signal sequence may be homologous (naturally occurring) or heterologous to the ChMIrp gene or cDNA. Additionally, a signal sequence may be chemically synthesized using methods set forth above. In most cases, secretion of an ChMIrp polypeptide from the host cell via the presence of a signal peptide will result in the removal of the signal peptide from the ChMIrp polypeptide.

The signal sequence may be a component of the vector, or it may be a part of ChMIrp DNA that is inserted into the vector. The native ChMIrp DNA encodes a signal sequence at the amino terminus of the protein that is cleaved during post-translational processing of the molecule to form the mature ChMIrp polypeptide product. Included within the scope of this invention are ChMIrp nucleotides with the native signal sequence as well as ChMIrp nucleotides wherein the native signal sequence is deleted and replaced with a heterologous signal sequence. The heterologous signal sequence selected should be one that is recognized and processed, *i.e.*, cleaved by a signal peptidase, by the host cell. For prokaryotic host cells that do not recognize and process the native ChMIrp signal sequence, the signal sequence is substituted by a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, or heat-stable enterotoxin II leaders. For yeast secretion, the native ChMIrp signal sequence may be substituted by the yeast invertase, alpha factor, or acid phosphatase signal sequences. For mammalian cell expression the native signal sequence of the ChMIrp polypeptides is satisfactory, although other mammalian signal sequences may be suitable.

In some cases, such as where glycosylation is desired in a eukaryotic host cell expression system, one may manipulate the various nucleic acid or polypeptide sequences to improve glycosylation or yield. For example, one may alter the peptidase cleavage site of a particular signal peptide, or add prosequences, which also may affect glycosylation. The final protein product may have in the -1 position (relative to the first amino acid of the mature protein) one or more additional amino acids incident to expression, which may not have been totally removed. For example, the final protein product may have one or two amino acids found in the peptidase cleavage site, attached to the N-terminus. Alternatively, use of some enzyme cleavage sites may result in a slightly truncated form of the desired ChMIrp polypeptide, if the enzyme cuts at an site area within the mature polypeptide.

In many cases, transcription of a nucleic acid molecule is increased by the presence of one or more introns in the vector; this is particularly true where a polypeptide is produced in eukaryotic host cells, especially mammalian host cells. The introns used may be naturally occurring within the ChMIrp gene, especially

where the gene used is a full length genomic sequence or a fragment thereof. Where the intron is not naturally occurring within the gene (as for most cDNAs), the intron(s) may be obtained from another source. The position of the intron with respect to 5′-flanking sequences and the ChMIrp gene is generally important, as the intron must be transcribed to be effective. Thus, when an ChMIrp cDNA molecule is being transcribed, the preferred position for the intron is 3′ to the transcription start site, and 5′ to the polyA transcription termination sequence. Preferably, the intron or introns will be located on one side or the other (*i.e.*, 5′ or 3′) of the cDNA such that it does not interrupt the coding sequence. Any intron from any source, including viral, prokaryotic and eukaryotic (plant or animal) organisms, may be used to practice this invention, provided that it is compatible with the host cell(s) into which it is inserted. Also included herein are synthetic introns. Optionally, more than one intron may be used in the vector.

The expression and cloning vectors of the present invention will typically contain a promoter that is recognized by the host organism and operably linked to the molecule encoding the ChMIrp polypeptide.

Promoters are untranscribed sequences located upstream (5′) to the start codon of a structural gene (generally within about 100 to 1000 bp) that control the transcription of the structural gene. Promoters are conventionally grouped into one of two classes, inducible promoters and constitutive promoters. Inducible promoters initiate increased levels of transcription from DNA under their control in response to some change in culture conditions, such as the presence or absence of a nutrient or a change in temperature. Constructive promoters, on the other hand, initiate continual gene product production; that is, there is little or no control over gene expression. A large number of promoters, recognized by a variety of potential host cells, are well known. A suitable promoter is operably linked to the DNA encoding ChMIrp polypeptide by removing the promoter from the source DNA by restriction enzyme digestion and inserting the desired promoter sequence into the vector. The native ChMIrp promoter sequence may be used to direct amplification and/or expression of ChMIrp nucleic acid molecule. A heterologous promoter is preferred, however, if it permits greater transcription and higher yields of the

expressed protein as compared to the native promoter, and if it is compatible with the host cell system that has been selected for use.

Promoters suitable for use with prokaryotic hosts include the beta-lactamase and lactose promoter systems; alkaline phosphatase, a tryptophan (trp) promoter system; and hybrid promoters such as the tac promoter. Other known bacterial promoters are also suitable. Their sequences have been published, thereby enabling one skilled in the art to ligate them to the desired DNA sequence(s), using linkers or adapters as needed to supply any useful restriction sites.

Suitable promoter for use with yeast hosts are also well known in the art. Yeast enhancers are advantageously used with yeast promoters. Suitable promoters for use with mammalian host cells are well known and include, but are not limited to, those obtained from the genomes of viruses such as polyoma virus, fowl pox virus, adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus, herpes virus and most preferably Simian Virus 40 (SV40). Other suitable mammalian promoters include heterologous mammalian promoters, *e.g.*, heat-shock promoters and the actin promoter.

Additional promoters which may be of interest in controlling ChMIrp gene transcription include, but are not limited to, the SV40 early promoter region (Bernoist and Chambon, *Nature*, 290: 304-310, 1981), the CMV promoter; the promoter contained in the 3′ long terminal repeat (LTR) of Rous sarcoma virus (RSV), (Yamamoto *et al.*, *Cell,* 22*:* 787-797, 1980), the herpes thymidine kinase promoter, (Wagner *et al., Proc. Natl. Acad. Sci. U.S.A.*, 78: 144-1445, 1981), the regulatory sequences of the metallothionine gene, (Brinster *et al., Nature*, 296: 39-42, 1982), prokaryotic expression vectors such as the beta-lactamase promoter, (Villa-Kamaroff *et al., Proc. Natl. Acad. Sci. U.S.A.*, 75: 3727-3731, 1978), or the tac promoter, (DeBoer *et al., Proc. Natl. Acad. Sci. U.S.A.*, 80: 21-25, 1983). Also of interest are the following animal transcriptional control regions, which exhibit tissue specificity and have been utilized in transgenic animals: the elastase I gene control region which is active in pancreatic acinar cells, (Swift *et al., Cell*, 38: 639-646, 1984; Ornitz *et al., Cold Spring Harbor Symp. Quant. Biol.,* 50:399-409,

1986; MacDonald, *Hepatology,* 7: 425-515 1987); the insulin gene control region which is active in pancreatic beta cells (Hanahan, *Nature,* 315: 115-122 1985); the immunoglobulin gene control region which is active in lymphoid cells (Grosschedl *et al., Cell*, 38: 647-658 1984; Adames *et al., Nature*, 318: 533-538, 1985; Alexander *et al., Mol. Cell. Biol.*, *7*: 1436-1444, 1987); the mouse mammary tumor virus control region which is active in testicular, breast, lymphoid and mast cells (Leder *et al., Cell*, 45*:* 485-495, 1986), the albumin gene control region which is active in liver (Pinkert *et al., Genes and Devel.*, 1: 268-276, 1987); the alphafetoprotein gene control region which is active in liver (Krumlauf *et al., Mol. Cell. Biol.*, 5: 1639-1648, 1985; Hammer *et al., Science*, 235: 53-58, 1987); the alpha 1-antitrypsin gene control region which is active in the liver (Kelsey *et al., Genes and Devel.*, 1: 161-171, 1987); the beta-globin gene control region which is active in myeloid cells (Mogram *et al., Nature*, 315: 338-340, 1985; Kollias *et al., Cell*, 46: 89-94, 1986); the myelin basic protein gene control region which is active in oligodendrocyte cells in the brain (Readhead *et al., Cell*, 48: 703-712, 1987); the myosin light chain-2 gene control region which is active in skeletal muscle (Sani, *Nature*, 314: 283-286, 1985); and the gonadotropic releasing hormone gene control region which is active in the hypothalamus (Mason *et al., Science*, 234: 1372-1378, 1986).

An enhancer sequence may be inserted into the vector to increase the transcription of a DNA encoding a ChMIrp polypeptide of the present invention by higher eukaryotes. Enhancers are cis-acting elements of DNA, usually about 10-300 bp in length, that act on the promoter to increase its transcription. Enhancers are relatively orientation and position independent. They have been found 5′ and 3′ to the transcription unit. Several enhancer sequences available from mammalian genes are known (*e.g.*, globin, elastase, albumin, alpha-feto-protein and insulin). Typically, however, an enhancer from a virus will be used. The SV40 enhancer, the cytomegalovirus early promoter enhancer, the polyoma enhancer, and adenovirus enhancers are exemplary enhancing elements for the activation or upregulation of eukaryotic promoters. While an enhancer may be spliced into the vector at a position 5′ or 3′ to ChMIrp nucleic acid molecule, it is typically located at a site 5′ from the promoter.

Expression vectors of the invention may be constructed from a starting vector such as a commercially available vector. Such vectors may or may not contain all of the desired flanking sequences. Where one or more of the desired flanking sequences are not already present in the vector to be used, they may be individually obtained and ligated into the vector. Methods used for obtaining each of the flanking sequences are well known to one skilled in the art.

Preferred vectors for practicing this invention are those which are compatible with bacterial, insect, and mammalian host cells. Such vectors include, *inter alia*, pCRII, pCR3, and pcDNA3.1 (Invitrogen Company, Carlsbad, CA), pBSII (Stratagene Company, La Jolla, CA), pET15 (Novagen, Madison, WI), pGEX (Pharmacia Biotech, Piscataway, NJ), pEGFP-N2 (Clontech, Palo Alto, CA), pETL (BlueBacII; Invitrogen), pDSR-alpha (PCT Publ. No. WO 90/14364), and pFastBac1, pFastBacHTand pFastBacDual (Gibco/BRL, Grand Island, NY).

Additional suitable vectors include, but are not limited to, cosmids, plasmids or modified viruses, but it will be appreciated that the vector system must be compatible with the selected host cell. Such vectors include, but are not limited to plasmids such as Bluescript® plasmid derivatives (a high copy number ColE1-based phagemid, Stratagene Cloning Systems Inc., La Jolla CA), PCR cloning plasmids designed for cloning Taq-amplified PCR products (e.g., TOPO™ TA Cloning® Kit, PCR2.1® plasmid derivatives, Invitrogen, Carlsbad, CA), and mammalian, yeast or virus vectors such as a baculovirus expression system (pBacPAK plasmid derivatives, Clontech, Palo Alto, CA).

After the vector has been constructed and a nucleic acid molecule encoding an ChMIrp polypeptide has been inserted into the proper site of the vector, the completed vector may be inserted into a suitable host cell for amplification and/or polypeptide expression. The transformation of an expression vector for a ChMIrp polypeptide into a selected host cell may be accomplished by well-known methods such as transfection, infection, calcium chloride, electroporation, microinjection, lipofection or the DEAE-dextran method or other known techniques. The method selected will in part be a function of the type of host cell to be used. These methods

and other suitable methods are well-known to the skilled artisan and are set forth, for example in Sambrook *et al., supra*.

Host cells may be prokaryotic host cells (such as *E. coli*) or eukaryotic host cells (such as a yeast cell, an insect cell, or a vertebrate cells). The host cell, when cultured under appropriate conditions, synthesizes a ChMIrp polypeptide which can subsequently be collected from the culture medium (if the host cell secretes it into the medium) or directly from the host cell producing it (if it is not secreted). The selection of an appropriate host cell will depend upon various factors, such as desired expression levels, polypeptide modifications that are desirable or necessary for activity, such as glycosylation or phosphorylation, and ease of folding into a biologically active molecule.

Yeast and mammalian cells are preferred hosts of the present invention. The use of such hosts provides substantial advantages in that they can also carry out post-translational peptide modifications including glycosylation. A number of recombinant DNA strategies exist which utilize strong promoter sequences and high copy number of plasmids which can be utilized for production of the desired proteins in these hosts.

Yeast recognize leader sequences on cloned mammalian gene products and secrete peptides bearing leader sequences (*i.e.*, pre-peptides). Mammalian cells provide post-translational modifications to protein molecules including correct folding or glycosylation at correct sites.

Mammalian cells which may be useful as hosts include cells of fibroblast origin such as VERO or CHO-K1, and their derivatives. For a mammalian host, several possible vector systems are available for the expression of the desired ChMIrp polypeptide. A wide variety of transcriptional and translational regulatory sequences may be employed, depending upon the nature of the host. The transcriptional and translational regulatory signals may be derived from viral sources, such as adenovirus, bovine papilloma virus, simian virus, or the like, where the regulatory signals are associated with a particular gene which has a high level of expression. Alternatively, promoters from mammalian expression products, such as actin, collagen, myosin, etc., may be employed. Transcriptional initiation regulatory

signals may be selected which allow for repression or activation, so that expression of the genes can be modulated. Useful signals are regulatory signals which are temperature-sensitive so that by varying the temperature, expression can be repressed or initiated, or are subject to chemical regulation, *e.g.*, metabolite.

As is widely known, translation of eukaryotic mRNA is initiated at the codon which encodes the first methionine. For this reason, it is preferable to ensure that the linkage between a eukaryotic promoter and a DNA sequence which encodes the desired receptor molecule does not contain any intervening codons which are capable of encoding a methionine (i.e., AUG). The presence of such codons results either in the formation of a fusion protein (if the AUG codon is in the same reading frame as the desired receptor molecule encoding DNA sequence) or a frame-shift mutation (if the AUG codon is not in the same reading frame as the desired ChMIrp polypeptide encoding sequence).

The expression of the ChMIrp polypeptides can also be accomplished in procaryotic cells. Preferred prokaryotic hosts include bacteria such as *E. coli*, Bacillus, Streptomyces, Pseudomonas, Salmonella, Serratia, etc. The most preferred prokaryotic host is *E. coli*. Bacterial hosts of particular interest include *E. coli* K12 strain 294 (ATCC 31446), *E. coli* X1776 (ATCC 31537), *E. coli* W3110 (F^-, lambda$^-$, prototrophic (ATCC 27325)), and other enterobacteria (such as *Salmonella typhimurium* or *Serratia marcescens*), and various Pseudomonas species. The prokaryotic host must be compatible with the replicon and control sequences in the expression plasmid.

To express the desired ChMIrp polypeptide in a prokaryotic cell (such as, for example, *E. coli*, *B. subtilis*, Pseudomonas, Streptomyces, etc.), it is necessary to operably link the desired receptor molecule encoding sequence to a functional prokaryotic promoter. Such promoters may be either constitutive or, more preferably, regulatable (*i.e.*, inducible or derepressible). Examples of constitutive promoters include the int promoter of bacteriophage λ, and the bla promoter of the β-lactamase gene of pBR322, etc. Examples of inducible prokaryotic promoters include the major right and left promoters, of bacteriophage λ (P_L and P_R), the trp, recA, lacZ, lacI, gal, and tac promoters of *E. coli*, the α-amylase (Ulmanen *et al.*, *J.*

Bacteriol., 162: 176-182, 1985), the σ-28-specific promoters of *B. subtilis* (Gilman *et al.*, *Gene*, 32: 11-20, 1984), the promoters of the bacteriophages of Bacillus (Gryczan, T. J., In: *The Molecular Biology of the Bacilli*, Academic Press, Inc., New York, NY, 1982), and Streptomyces promoters (Ward *et al.*, *Mol. Gen. Genet.*, 203: 468-478, 1986). Prokaryotic promoters are reviewed by Glick, B.R., *J. Ind. Microbiol,* 1:2 77-282, 1987; Cenatiempo, *Biochimie,* 68: 505-516, 1986; and Gottesman, *Ann. Rev. Genet.*, 18: 415-442, 1984.

Proper expression in a prokaryotic cell also requires the presence of a ribosome binding site upstream from the gene-encoding sequence. Such ribosome binding sites are disclosed, for example, by Gold *et al.* (*Ann. Rev. Microbiol.*, 35: 365-404, 1981).

The desired ChMIrp polypeptide encoding sequence and an operably linked promoter may be introduced into a recipient prokaryotic or eukaryotic cell either as a non-replicating DNA (or RNA) molecule, which may either be linear or, more preferably, a closed covalent circular molecule. Since such molecules are incapable of autonomous replication, the expression of the desired receptor molecule may occur through the transient expression of the introduced sequence. Alternatively, permanent expression may occur through the integration of the introduced sequence into the host chromosome.

In one embodiment, a vector is employed which is capable of integrating the desired gene sequences into the host cell chromosome. Cells which have stably integrated the introduced DNA into their chromosomes can be selected by also introducing one or more markers which allow for selection of host cells which contain the expression vector. The marker may complement an auxotrophy in the host (such as leu2l, or ura3, which are common yeast auxotrophic markers), biocide resistance, *e.g.*, antibiotics, or heavy metals, such as copper, or the like. The selectable marker gene can either be directly linked to the DNA gene sequences to be expressed, or introduced into the same cell by co-transfection.

In a preferred embodiment, the introduced sequence will be incorporated into a plasmid or viral vector capable of autonomous replication in the recipient host. Any of a wide variety of vectors may be employed for this purpose.

Factors of importance in selecting a particular plasmid or viral vector include, for *e.g.* the ease with which recipient cells that contain the vector may be recognized and selected from those recipient cells which do not contain the vector; the number of copies of the vector which are desired in a particular host; and whether it is desirable to be able to "shuttle" the vector between host cells of different species.

Any of a series of yeast gene expression systems can also be utilized. Examples of such expression vectors include the yeast 2-micron circle, the expression plasmids YEP13, YVP and YRP, etc., or their derivatives. Such plasmids are well known in the art (Botstein *et al.*, *Miami Wntr. Symp.*, 19: 265-274, 1982; Broach, J. R., In: *The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., p. 445-470, 1981; Broach, *Cell*, 28: 203-204, 1982).

For a mammalian host, several possible vector systems are available for expression. One class of vectors utilize DNA elements which provide autonomously replicating extra-chromosomal plasmids, derived from animal viruses such as bovine papilloma virus, polyoma virus, adenovirus, or SV40 virus. A second class of vectors relies upon the integration of the desired gene sequences into the host chromosome. Cells which have stably integrated the introduced DNA into their chromosomes may be selected by also introducing one or more markers which allow selection of host cells which contain the expression vector. The marker may provide for prototropy to an auxotrophic host, biocide resistance, *e.g.*, antibiotics, or heavy metals, such as copper or the like. The selectable marker gene can either be directly linked to the DNA sequences to be expressed, or introduced into the same cell by co-transformation. Additional elements may also be needed for optimal synthesis of mRNA. These elements may include splice signals, as well as transcription promoters, enhancers, and termination signals. The cDNA expression vectors incorporating such elements include those described by Okayama, H. (*Mol. Cell. Biol.* 3: 280, 1983). Preferred eukaryotic vectors include PWLNEO, PSV2CAT, POG44, PXTl, pSG, pSVK3, pBPV, pMSG, pSVL (Pharmacia).

Preferred prokaryotic vectors include plasmids such as those capable of replication in *E. coli* such as, for example, pBR322, ColE1, pSC101, pACYC 184,

πVX, pQE70, pQE60, pQE9, pBG, pD10, Phage script, psix174, pbmescript SK, pbsks, pNH8A, pNHIBa, pNH18A, pNH46A (SL rare gone), ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5. Such plasmids are, for example, disclosed by Maniatis, *et al.* (In: *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, NY, 1982). Bacillus plasmids include pC194, pC221, pT127, etc. Such plasmids are disclosed by Gryczan (In: *The Molecular Biology of the Bacilli*, Academic Press, New York, NY, pp. 307-329, 1982). Suitable Streptomyces plasmids include pISJ101 (Kendall *et al.*, *J. Bacteriol.*, 169: 4177-4183, 1987), and Streptomyces bacteriophages such as φC31 (Chater *et al.*, In: *Sixth International Symposium on Actinomycetales Biology*, Akademiai Kaidi, Budapest, Hungary, pp 45-541, 1986). Pseudomonas plasmids are reviewed by John *et a. (Rev. Infect. Dis.*, 8: 693-704, 1986) and Izaki (*Jpn. J. Bacteriol.*, 33: 729-742, 1978). However, any other plasmid or vector may be used as long as they are replicable and viable in the host cell.

Once the vector or DNA sequence containing the constructs has been prepared for expression, the DNA constructs may be introduced into an appropriate host. Various techniques may be employed, such as a protoplast fusion, calcium phosphate precipitation, electroporation or other conventional techniques. After the fusion, the cells are grown in media and screened for appropriate activities. Expression of the sequence results in the production of the ChMIrp polypeptide.

Suitable host cells or cell lines may be mammalian cells, such as Chinese hamster ovary cells (CHO) (ATCC No. CCL61), CHO DHFR-cells (Urlaub *et al., Proc. natl. Acad. Sci. U.S.A, 97*: 4216-4220, 1980) human embryonic kidney cells (HEK) 293 or 293T cells (ATCC No. CRL1573), 3T3 cells (ATCC No. CCL92). The selection of suitable mammalian host cells and methods for transformation, culture, amplification, screening and product production and purification are known in the art. Other suitable mammalian cell lines, are the monkey COS-1 (ATCC No. CRL1650)and COS-7 (ATCC No. CEL1651) cell lines, and the CV-1 cell line (ATCC No. CCL70). Further exemplary mammalian host cells include primate cell lines and rodent cell lines, including transformed cell lines. Normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, as

well as primary explants, are also suitable. Candidate cells may be genotypically deficient in the selection gene, or may contain a dominantly acting selection gene. Other suitable mammalian cell lines include, but are not limited to, mouse neuroblastoma N2A cells, HeLa cells, mouse L-929 cells, 3T3 lines derived form Swiss, Balb-c or NIH mice, BHK or HaK hamster cell lines, which are also available from the ATCC. Each of these cell lines is known by and available to those skilled in the art of protein expression..

Similarly useful as host cells suitable for the present invention are bacterial cells. For example, the various strains of *E. coli* (*e.g.*, HB101 (ATCC No. 33694), DH5α, DH10, and MC1061(ATCC No. 53338)) are well-known as host cells in the field of biotechnology. Various strains of *B. subtilis*, *Pseudomonas spp.*, other *Bacillus spp.*, *Streptomyces spp.*, and the like may also be employed in this method.

Many strains of yeast cells known to those skilled in the art are also available as host cells for expression of the polypeptides of the present invention (*e.g.* Saccharomyces, Pichia, Candida, Hansenula, and Torulopsis) (Bitter, G., *Heterologous Gene Expression in Yeast* In: Berger, S.L. and Kimmel, A.R., 152: 673-684, 1987). Preferred yeast strains include, for example, *Saccharomyces cerevisiae,* which can be transformed readily with DNA either by preparation of spheroplasts or by treatment with alkaline salts such as LiCl. (Itoh *et al., J. Bacteriol.,* 153: 163, 1983). Some proteins expressed in yeast cells are efficiently secreted into the culture medium while others accumulate intracellularly.

Additionally, insect cell systems may be utilized in the methods of the present invention. Such systems are described for example in Kitts *et al. Biotechniques*, 14: 810-817, 1993); Lucklow *et al.* (*Curr. Opin. Biotechnol.*, 4: 564-572, 1993); and Lucklow *et al (J. Virol.*, 67: 4566-4579, 1993). Preferred insect cells are Sf-9 and Hi5 (Invitrogen, Carlsbad, CA). Preferably, the insect cells are infected with recombinant baculovirus particles which have the ChMIrp polynucleotide sequence incorporated into its genome. The baculovirus infection leads to efficient eukaryotic expression of recombinant ChMIrp. Such baculovirus expression systems include the Bac-to-Bac® System (Life Technologies) and the BacPAK™ System (Clontech).

Polypeptide Production

Host cells comprising a ChMIrp polypeptide expression vector may be cultured using standard media well known to the skilled artisan. The media will usually contain all nutrients necessary for the growth and survival of the cells. Suitable media for culturing *E. coli* cells are for example, Luria Broth (LB) and/or Terrific Broth (TB). Suitable media for culturing eukaryotic cells include, Roswell Park Memorial Institute medium 1640 (RPMI 1640), Minimal Essential Medium (MEM) and/or Dulbecco's Modified Eagles Medium (DMEM), all of which may be supplemented with serum and/or growth factors as indicated by the particular cell line being cultured. A suitable medium for insect cultures is Grace's medium supplemented with yeastolate, lactalbumin hydrolysate, and/or fetal calf serum as necessary.

Typically, an antibiotic or other compound useful for selective growth of transformed cells is added as a supplement to the media. The compound to be used will be dictated by the selectable marker element present on the plasmid with which the host cell was transformed. For example, where the selectable marker element is kanamycin resistance, the compound added to the culture medium will be kanamycin. Other compounds for selective growth include ampicillin, tetracycline, and neomycin.

The amount of ChMIrp polypeptide produced by a host cell can be evaluated using standard methods known in the art. Such methods include, without limitation, Western blot analysis, SDS-polyacrylamide gel electrophoresis, non-denaturing gel electrophoresis, HPLC separation, immunoprecipitation, and/or activity assays.

If a ChMIrp polypeptide has been designed to be secreted from the host cells, the majority of polypeptide may be found in the cell culture medium. If however, the ChMIrp polypeptide is not secreted from the host cells, it will be present in the cytoplasm and/or the nucleus (for eukaryotic host cells) or in the cytosol (for bacterial host cells).

For a ChMIrp polypeptide situated in the host cell cytoplasm and/or nucleus (for eukaryotic host cells) or in the cytosol (for bacterial host cells), the host cells are typically first disrupted mechanically or with detergent to release the intra-cellular contents into a buffered solution. The ChMIrp polypeptide can then be isolated from this solution.

Purification of a ChMIrp polypeptide from solution can be accomplished using a variety of techniques. If the polypeptide has been synthesized such that it contains a tag such as Hexahistidine (ChMIrp polypeptide/hexaHis) or other small peptide such as FLAG (Eastman Kodak Co., New Haven, CT) or myc (Invitrogen, Carlsbad, CA) or the IgG Fc fragment fused at either its carboxyl or amino terminus, it may essentially be purified in a one-step process by passing the solution through an affinity column where the column matrix has a high affinity for the tag or for the polypeptide directly (*i.e.*, a monoclonal antibody specifically recognizing ChMIrp polypeptide). For example, polyhistidine binds with great affinity and specificity to nickel, thus an affinity column of nickel (such as the Qiagen® nickel columns) can be used for purification of ChMIrp polypeptide/polyHis. (*See* for example, Ausubel *et al., eds.*, *Current Protocols in Molecular Biology*, Section 10.11.8, John Wiley & Sons, New York, NY, 1993).

Additionally, the ChMIrp polypeptide may be purified through the use of a monoclonal antibody which is capable of specifically recognizing and binding to the ChMIrp polypeptide.

Where a ChMIrp polypeptide is prepared without a tag attached, and no antibodies are available, other well known procedures for purification can be used. Such procedures include, without limitation, ion exchange chromatography, molecular sieve chromatography, HPLC, native gel electrophoresis in combination with gel elution, and preparative isoelectric focusing ("Isoprime" machine/technique, Hoefer Scientific). In some cases, two or more of these techniques may be combined to achieve increased purity.

If a ChMIrp polypeptide is produced intracellularly, the intracellular material (including inclusion bodies for gram-negative bacteria) can be extracted from the host cell using any standard technique known to the skilled artisan. For example,

the host cells can be lysed to release the contents of the periplasm/cytoplasm by detergent lysis (*e.g.*, Triton X-100), by French press, homogenization, and/or sonication followed by centrifugation.

If a ChMIrp polypeptide has formed inclusion bodies in the cytosol, the inclusion bodies can often bind to the inner and/or outer cellular membranes and thus will be found primarily in the pellet material after centrifugation. The pellet material can then be treated at pH extremes or with a chaotropic agent such as a detergent, guanidine, guanidine derivatives, urea, or urea derivatives in the presence of a reducing agent such as dithiothreitol at alkaline pH or tris carboxyethyl phosphine at acid pH to release, break apart, and solubilize the inclusion bodies. The ChMIrp polypeptide in its now soluble form can then be analyzed using gel electrophoresis, immunoprecipitation or the like. If it is desired to isolate the ChMIrp polypeptide, isolation may be accomplished using standard methods such as those described herein and in Marston *et al.* (*Meth. Enz.*, 182: 264-275, 1990).

In some cases, a ChMIrp polypeptide may not be biologically active upon isolation. Various methods for "refolding" or converting the polypeptide to its tertiary structure and generating disulfide linkages, can be used to restore biological activity. Such methods include exposing the solubilized polypeptide to a pH usually above 7 and in the presence of a particular concentration of a chaotrope. The selection of chaotrope is very similar to the choices used for inclusion body solubilization, but usually the chaotrope is used at a lower concentration and is not necessarily the same as chaotropes used for the solubilization. In most cases the refolding/oxidation solution will also contain a reducing agent or the reducing agent plus its oxidized form in a specific ratio to generate a particular redox potential allowing for disulfide shuffling to occur in the formation of the protein's cysteine bridge(s). Some of the commonly used redox couples include cysteine/cystamine, glutathione (GSH)/dithiobis GSH, cupric chloride, dithiothreitol (DTT)/dithiane DTT,

2-2-mercaptoethanol (bME)/dithio-b(ME). A cosolvent may be used to increase the efficiency of the refolding, and the more common reagents used for this purpose

include glycerol, polyethylene glycol of various molecular weights, arginine and the like.

If inclusion bodies are not formed to a significant degree upon expression of a ChMIrp polypeptide, then the polypeptide will be found primarily in the supernatant after centrifugation of the cell homogenate. The polypeptide may be further isolated from the supernatant using methods such as those described herein.

In situations where it is preferable to partially or completely purify a ChMIrp polypeptide such that it is partially or substantially free of contaminants, standard methods known to the one skilled in the art may be used. Such methods include, without limitation, separation by electrophoresis followed by electroelution, various types of chromatography (affinity, immunoaffinity, molecular sieve, and/or ion exchange), and/or high pressure liquid chromatography. In some cases, it may be preferable to use more than one of these methods for complete purification.

ChMIrp polypeptides, fragments, and/or derivatives thereof may also be prepared by chemical synthesis methods (such as solid phase peptide synthesis) using techniques known in the art such as those set forth by Merrifield *et al.*(*J. Am. Chem. Soc.*, 85: 2149, 1963); Houghten *et al.* (*Proc Natl Acad. Sci. USA*, 82: 5132, 1985); and Stewart and Young (*Solid Phase Peptide Synthesis*, Pierce Chemical Co., Rockford, IL, 1984). Such polypeptides may be synthesized with or without a methionine on the amino terminus. Chemically synthesized ChMIrp polypeptides or fragments may be oxidized using methods set forth in these references to form disulfide bridges. ChMIrp polypeptides, fragments or derivatives are expected to have comparable biological activity to the corresponding ChMIrp polypeptides, fragments or derivatives produced recombinantly or purified from natural sources, and thus may be used interchangeably with recombinant or natural ChMIrp polypeptide.

Another means of obtaining ChMIrp polypeptide is via purification from biological samples such as source tissues and/or fluids in which the ChMIrp polypeptide is naturally found. Such purification can be conducted using methods for protein purification as described above. The presence of the ChMIrp polypeptide

during purification may be monitored using, for example, an antibody prepared against recombinantly produced ChMIrp polypeptide or peptide fragments thereof.

A number of additional methods for producing nucleic acids and polypeptides are known in the art, and the methods can be used to produce polypeptides having specificity for ChMIrp. *See* for example, Roberts *et al., Proc. Natl. Acad. Sci U.S.A., 94*:12297-12303, 1997, which describes the production of fusion proteins between an mRNA and its encoded peptide. See also Roberts, R., *Curr. Opin. Chem. Biol., 3*:268-273, 1999. Additionally, U.S. Patent No. 5,824,469 describes methods of obtaining oligonucleotides capable of carrying out a specific biological function. The procedure involves generating a heterogeneous pool of oligonucleotides, each having a 5′ randomized sequence, a central preselected sequence, and a 3′ randomized sequence. The resulting heterogeneous pool is introduced into a population of cells that do not exhibit the desired biological function. Subpopulations of the cells are then screened for those which exhibit a predetermined biological function. From that subpopulation, oligonucleotides capable of carrying out the desired biological function are isolated.

U.S. Patent Nos. 5,763,192; 5,814,476; 5,723,323; and 5,817,483 describe processes for producing peptides or polypeptides. This is done by producing stochastic genes or fragments thereof, and then introducing these genes into host cells which produce one or more proteins encoded by the stochastic genes. The host cells are then screened to identify those clones producing peptides or polypeptides having the desired activity.

Another method for producing peptides or polypeptides is described in PCT/US98/20094 (WO99/15650) filed by Athersys, Inc. Known as "Random Activation of Gene Expression for Gene Discovery" (RAGE-GD), the process involves the activation of endogenous gene expression or over-expression of a gene by in situ recombination methods. For example, expression of an endogenous gene is activated or increased by integrating a regulatory sequence into the target cell which is capable of activating expression of the gene by non-homologous or illegitimate recombination. The target DNA is first subjected to radiation, and a genetic promoter inserted. The promoter eventually locates a break at the front of a gene,

initiating transcription of the gene. This results in expression of the desired peptide or polypeptide.

It will be appreciated that these methods can also be used to create comprehensive ChMIrp protein expression libraries, which can subsequently be used for high throughput phenotypic screening in a variety of assays, such as biochemical assays, cellular assays, and whole organism assays (*e.g.*, plant, mouse, etc.).

Proteins, Polypeptides, Fragments, Variants and Muteins of ChMIrp:

Polypeptides of the invention include isolated ChMIrp polypeptides and polypeptides related thereto including fragments, variants, fusion polypeptides, and derivatives as defined hereinabove.

ChMIrp fragments of the invention may result from truncations at the amino terminus (with or without a leader sequence), truncations at the carboxy terminus, and/or deletions internal to the polypeptide. Most deletions and insertions, and substitutions in particular, are not expected to produce radical changes in the characteristics of the ChMIrp polypeptide. However, when it is difficult to predict the exact effect of the substitution, deletion, or insertion in advance of doing so, one skilled in the art will appreciate that the effect will be evaluated by routine screening assays. For example, a variant typically is made by site-specific mutagenesis of the ChMIrp polypeptide-encoding nucleic acid, expression of the variant nucleic acid in recombinant cell culture, and, optionally, purification from the cell culture, for example, by immunoaffinity adsorption on a polyclonal anti-ChMIrp antibody column (to absorb the variant by binding it to at least one remaining immune epitope). In preferred embodiments, truncations and/or deletions comprise about 10 amino acids, or about 20 amino acid, or about 50 amino acids, or about 75 amino acids, or about 100 amino acids, or more than about 100 amino acids. The polypeptide fragments so produced will comprise about 25 contiguous amino acids, or about 50 amino acids, or about 75 amino acids, or about 100 amino acids, or about 150 amino acids, or about 200 amino acids or about 250 amino acids, or about 275 amino acids, or 300 amino acids. Such ChMIrp polypeptides fragments may optionally comprise an amino terminal methionine residue.

ChMIrp polypeptide variants of the invention include one or more amino acid substitutions, additions and/or deletions as compared to SEQ ID NO: 2. In preferred embodiments, the variants have from 1 to 3, or from 1 to 5, or from 1 to 10, or from 1 to 15, or from 1 to 20, or from 1 to 25, or from 1 to 50, or from 1 to 75, or from 1 to 100, or more than 100 amino acid substitutions, insertions, additions and/or deletions, wherein the substitutions may be conservative, as defined above, or non-conservative or any combination thereof. More particularly, ChMIrp variants may comprise the amino acid sequence set out as SEQ ID NO: 2, wherein one or more amino acids from the group consisting of amino acids 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, or 317 is/are substituted with another amino acid. The variants may have additions of amino acid residues either at the carboxy terminus or at the amino terminus (with or without a leader sequence).

Preferred ChMIrp polypeptide variants include glycosylation variants wherein the number and/or type of glycosylation sites has been altered compared to native ChMIrp polypeptide. In one embodiment, ChMIrp variants comprise a greater or a lesser number of N-linked glycosylation sites. An N-linked glycosylation site is characterized by the sequence: Asn-X-Ser or Thr, where the amino acid residue designated as X may be any type of amino acid except proline. Substitution(s) of amino acid residues to create this sequence provides a potential new site for addition of an N-linked carbohydrate chain. Alternatively, substitutions to eliminate this sequence will remove an existing N-linked carbohydrate chain. Also provided is a rearrangement of N-linked carbohydrate chains wherein one or more N-linked glycosylation sites (typically those that are naturally occurring) are eliminated and one or more new N-linked sites are created.

One skilled in the art will be able to determine suitable variants of the native ChMIrp polypeptide using well known techniques. For example, one may be able to predict suitable areas of the molecule that may be changed without destroying biological activity. Also, one skilled in the art will realize that even areas that may be important for biological activity or for structure may be subject to conservative amino acid substitutions without destroying the biological activity or without adversely affecting the polypeptide structure.

For predicting suitable areas of the molecule that may be changed without destroying activity, one skilled in the art may target areas not believed to be important for activity. For example, when similar polypeptides with similar activities from the same species or from other species are known, one skilled in the art may compare the amino acid sequence of ChMIrp polypeptide to such similar polypeptides. After making such a comparison, one skilled in the art would be able to determine residues and portions of the molecules that are conserved among similar polypeptides. One skilled in the art would know that changes in areas of the ChMIrp molecule that are not conserved would be less likely to adversely affect biological activity and/or structure. One skilled in the art would also know that, even in relatively conserved regions, one could have likely substituted chemically similar

amino acids for the naturally occurring residues while retaining activity (*e.g.* conservative amino acid residue substitutions).

Also, one skilled in the art may review structure-function studies identifying residues in similar polypeptides that are important for activity or structure. In view of such a comparison, one skilled in the art can predict the importance of amino acid residues in ChMIrp that correspond to amino acid residues that are important for activity or structure in similar polypeptides. One skilled in the art may opt for chemically similar amino acid substitutions for such predicted important amino acid residues of ChMIrp .

If available, one skilled in the art can also analyze the crystal structure and amino acid sequence in relation to that structure in similar polypeptides. In view of that information, one skilled in the art may be able to predict the alignment of amino acid residues of ChMIrp polypeptide with respect to its three dimensional structure. One skilled in the art may choose not to make radical changes to amino acid residues predicted to be on the surface of the protein, since such residues may be involved in important interactions with other molecules.

Moreover, one skilled in the art could generate test variants containing a single amino acid substitution at each amino acid residue. The variants could be screened using activity assays disclosed in this application. Such variants could be used to gather information about suitable variants. For example, if one discovered that a change to a particular amino acid residue resulted in destroyed activity, variants with such a change would be avoided. In other words, based on information gathered from such experiments, when attempting to find additional acceptable variants, one skilled in the art would have known the amino acids where further substitutions should be avoided either alone or in combination with other mutations.

ChMIrp fusion polypeptides of the invention comprise ChMIrp polypeptides, fragments, variants, or derivatives fused to a heterologous functional portion of peptide(s) or protein(s). Heterologous peptide(s) and protein(s) include, but are not limited to, an epitope to allow for detection and/or isolation of a ChMIrp fusion polypeptide, a transmembrane receptor protein or a portion thereof, such as an extracellular domain, or a transmembrane, a ligand or a portion thereof which binds

to a transmembrane receptor protein, an enzyme or portion thereof which is catalytically active, a protein or peptide which promotes oligomerization, such as leucine zipper domain, and a protein or peptide which increase stability, or circulatory half-life, such as an immunoglobulin constant region. A ChMIrp polypeptide may be fused to itself or to a fragment, variant, or derivative thereof. Fusions may be made either at the amino terminus or at the carboxy terminus of a ChMIrp polypeptide, and may be direct with no linker or adapter molecule or may be through a linker or adapter molecule, such as one or more amino acid residues up to about 20 amino acids residues, or up to about 50 amino acid residues. A linker or adapter molecule may also be designed with a cleavage site for a DNA restriction endonuclease or for proteolytic cleavage to allow for separation and subsequent folding of the fused moieties.

Also envisioned as a part of the invention are circularly permuted structural analogs of the ChMIrp polypeptide.

The development of recombinant DNA methods has made it possible to study the effects of sequence transposition on protein folding, structure and function. The approach used in creating new sequences resembles that of naturally occurring pairs of proteins that are related by linear reorganization of their amino acid sequences (Cunningham *et al., Proc. Natl. Acad. Sci. U.S.A.,* 76: 3218-3222, 1979; Teather & Erfle, *J. Bacteriol.*, 172: 3837-3841, 1990; Schimming *et al.*, *Eur. J. Biochem.,* 204: 13-19, 1992; Yamiuchi and Minamikawa, *FEBS Lett.,* 260: 127-130, 1991; MacGregor *et al.*, *FEBS Lett.,* 378: 263-266, 1996). The first *in vitro* application of this type of rearrangement to proteins was described by Goldenberg and Creighton (*J. Mol. Biol.,* 165: 407-413, 1983). A new N-terminus is selected at an internal site (breakpoint) of the original sequence, the new sequence having the same order of amino acids as the original from the breakpoint until it reaches an amino acid that is at or near the original C-terminus. At this point the new sequence is joined, either directly or through an additional portion of sequence (linker), to an amino acid that is at or near the original N-terminus, and the new sequence continues with the same sequence as the original until it reaches a point that is at or near the

Cysteinyl residues most commonly are reacted with α-haloacetates (and corresponding amines), such as chloroacetic acid or chloroacetamide, to give carboxymethyl or carbocyamidomethyl derivatives. Cysteinyl residues also are derivatized by reaction with bromotrifluoroacetone, α-bromo-β(5-imidozoyl)propionic acid, chloroacetyl phosphate, N-alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloromercuribenzoate, 2-chloromercuri-4-nitrophenol, orchloro-7-nitrobenzo-2-oxa-1,3-diazole.

Histidyl residues are derivatized by reaction with diethylprocarbonate at pH 5.5-7.0 because this agent is relatively specific for the histidyl side chain. Para-bromophenacyl bromide also is useful; the reaction is preferably performed in 0.1 M sodium cacodylate at pH 6.0.

Lysinyl and amino terminal residues are reacted with succinic or carboxylic acid anhydrides. Derivatization with these agents has the effect of reversing the charge of the lysinyl residues. Other suitable reagents for derivatizing α-amino-containing residues include imidoesters such as methyl picolinimidate; pyridoxal phosphate; pyridoxal; chloroborohydride; trinitrobenzenesulfonic acid; O-methylissurea; 2,4 pentanedione; and transaminase catalyzed reaction with glyoxylate.

Arginyl residues are modified by reaction with one or several conventional reagents, among them phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and ninhydrin. Derivatization of arginine residues requires that the reaction be performed in alkaline conditions because of the high pK_a of the guanidine functional group. Furthermore, these reagents may react with the groups of lysine as well as the arginine Epsilon-amino group.

The specific modification of tyrosyl residues *per se* has been studied extensively, with particular interest in introducing spectral labels into tyrosyl residues by reaction with aromatic diazonium compounds or tetranitromethane. Most commonly, N-acetylimidizol and tetranitromethane are used to form O-acetyl tyrosyl species and 3-nitro derivatives, respectively. Tyrosyl residues are iodinated using ^{125}I or ^{131}I to prepare labeled proteins for use in radioimmunoassay, the chloramine method described above being suitable.

Carboxyl side groups (aspartyl or glutamyl) are selectively modified by reaction with carbodiimides (R^1) such as 1-cyclohexyl-3-(2-morpholinyl-(4-ethyl) carbodiimide or 1-ethyl-3 (4 azonia 4,4-dimethylpentyl) carbodiimide. Furthermore, aspartyl and glutamyl residues are converted to asparaginyl and glutaminyl residues by reaction with ammonium ions.

Derivatization with bifunctional agents is useful for crosslinking the ChMIrp polypeptides to water-insoluble support matrixes or surfaces for use in the method for cleaving the ChMIrp polypeptide-fusion polypeptide to release and recover the cleaved polypeptide. Commonly used crosslinking agents include, *e.g.*, 1,1-bis(diazoacetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homo-bifunctional imidoesters, including disuccinimidyl esters such as 3,3′-dithiiobis(succinimidylpropionate), and bifunctional maleimides such as bix-N-maleimido-1,8-octane. Derivatizing agents such as methyl-3-[p-azidophenyl dithio]propioimidate yield photoactivatable intermediates that are capable of forming cross links in the presence of light. Alternatively, reactive water-insoluble matrices such as cyanogen bromide-activated carbohydrates and the reactive substrates described in U.S. Pat. Nos. 3,969,287; 3,691,016; 4,195,128; 4,247,642; 4,229,537; and 4,330,440, are employed for protein immobilization.

Glutaminyl and asparaginyl residues are frequently deamidated to the corresponding glutamyl and aspartyl residues. Alternatively, these residues are deamidated under mildly acidic conditions. Either form of these residues falls within the scope of this invention.

Other modifications include hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or theonyl residues, methylation of the α-amino groups of lysine, arginine, and histidine side chains (Creighton,T. E. *Proteins: Structure and Molecule Properties*, W. H. Freeman & Co., San Francisco, pp. 79-86, 1983), acetylation of the N-terminal amine, and, in some instances, amidation of the C-terminal carboxyl groups. Such derivatives are chemically modified ChMIrp polypeptide compositions in which ChMIrp polypeptide is linked to a polymer. The polymer selected is typically water soluble so that the protein to

which it is attached does not precipitate in an aqueous environment, such as a physiological environment. The polymer selected is usually modified to have a single reactive group, such as an active ester for acylation or an aldehyde for alkylation, so that the degree of polymerization may be controlled as provided for in the present methods.

The polymer may be of any molecular weight, and may be branched or unbranched. The polymers each typically have an average molecular weight of between about 2 kDa to about 100 kDa (ther term "about" indicating that in preparations of water soluble polymer, some molecules will weigh more, some less, than the stated molecular weight). The average molecular weight of each polymer is preferably between about 5 kDa and about 50 kDa, more preferably between about 12 kDa and about 40 kDa and most preferably between about 20 kDa to about 35 kDa. Suitable water soluble polymers or mixtures thereof include but are not limited to N-linked or O-linked carbohydrates; sugars; phosphates; polyethylene glycol (PEG) (including the forms of PEG that have been used to derivatize proteins, including mono-(C1-C10) alkoxy- or aryloxy-polyethylene glycol), monomethoxy-polyethylene glycol; dextran (such as low molecular weight dextran of, for example about 6 kD;, cellulose, or other carbohydrat- based polymers, poly-(N-vinyl pyrrolidone) polyethylene glycol, propylene glycol homopolymers, a polypropylene oxide/ethylene oxide co-polymer, polyoxyethylated polyols (*e.g.*, glycerol) and polyvinyl alcohol. Also encompassed by the present invention are bifunctional crosslinking molecules which may be used to prepare covalently attached multimers of the polypeptide comprising the amino acid sequence of SEQ ID NO: 2 or a ChMIrp polypeptide variant.

In general, chemical derivatization may be performed under any suitable condition used to react a protein with an activated polymer molecule. Methods for preparing chemical derivatives of polypeptides will generally comprise the steps of (a) reacting the polypeptide with the activated polymer molecule (such as a reactive ester or aldehyde derivative of the polymer molecule) under conditions whereby the polypeptide comprising the amino acid sequence of SEQ ID NO: 2, or a ChMIrp polypeptide variant becomes attached to one or more polymer molecules, and

(b) obtaining the reaction product(s). The optimal reaction conditions will be determined based on known parameters and the desired result. For example, the larger the ratio of polymer molecules:protein, the greater the percentage of attached polymer molecule. In one embodiment, the ChMIrp polypeptide derivative may have a single polymer molecule moiety at the amino terminus. (*See*, for example, U.S. Patent No. 5,234,784).

Pegylation of ChMIrp polypeptides may be carried out by any of the pegylation reactions known in the art, as described for example in the following references: *Focus on Growth Factors*, 3: 4-10, 1992; EP 0 154 316; and EP 0 401 384. Preferably, the pegylation is carried out via an acylation reaction or an alkylation reaction with a reactive polyethylene glycol molecule (or an analogous reactive water-soluble polymer) as described below.

A particularly preferred water-soluble polymer for use herein is polyethylene glycol, abbreviated PEG. As used herein, polyethylene glycol is meant to encompass any of the forms of PEG that have been used to derivatize other proteins, such as mono-(C1-C10) alkoxy- or aryloxy-polyethylene glycol. PEG is a linear or branched neutral polyether, available in a broad range of molecular weights, and is soluble in water and most organic solvants. PEG is effective at excluding other polymers or peptides when present in water, primarily through its high dynamic chain mobility and hydrophibic nature, thus creating a water shell or hydration sphere when attached to other proteins or polymer surfaces. PEG is nontoxic, non-immunogenic, and approved by the Food and Drug Administration for internal consumption.

Proteins or enzymes when conjugated to PEG have demonstrated bioactivity, non-antigenic properties, and decreased clearance rates when administered in animals (Veronese *et al., Preparation and Properties of Monomethoxypoly(ethylene glyco.)-modified Enzymes for Therapeutic Applications,* In: J. M. Harris *ed., Poly(Ethylene Clycol) Chemistry--Biotechnical and Biomedical Applications,* pp.127-36, 1992). This is due to the exclusion properties of PEG in preventing recognition by the immune system. In addition, PEG has been widely used in surface modification procedures to decrease protein adsorption and improve blood compatibility (Kim *et al., Ann. N.Y. Acad. Sci.* 516: 116-30, 1987; Jacobs *et*

al., Artif. Organs 12: 500-501, 1988; Park *et al., J. Poly. Sci*, 29(Part A):1 725-31, 1991). Hydrophobic polymer surfaces, such as polyurethanes and polystyrene were modified by the grafting of PEG (MW 3,400) and employed as nonthrombogenic surfaces. In these studies, surface properties (contact angle) were more consistent with hydrophilic surfaces, due to the hydrating effect of PEG. More importantly, protein (albumin and other plasma proteins) adsorption was greatly reduced, resulting from the high chain motility, hydration sphere, and protein exclusion properties of PEG.

PEG (MW 3,4000) was determined as an optimal size in surface immobilization studies (Park *et al., J. Biomed. Mat. Res.*, 26: 739-45, 1992), while PEG (MW 5,000) was most beneficial in decreasing protein antigenicity (Veronese *et al.*, In J. M. Harris. *et.al. eds*, *Poly(Ethylene Glycol) Chemistry--Biotechnical and Biomedical Applications* 127-36, *supra.*).

In general, chemical derivatization may be performed under any suitable conditions used to react a biologically active substance with an activated polymer molecule. Methods for preparing pegylated ChMIrp polypeptides will generally comprise the steps of (a) reacting the polypeptide with polyethylene glycol (such as a reactive ester or aldehyde derivative of PEG) under conditions whereby ChMIrp polypeptide becomes attached to one or more PEG groups, and (b) obtaining the reaction product(s). In general, the optimal reaction conditions for the acylation reactions will be determined based on known parameters and the desired result. For example, the larger the ratio of PEG: protein, the greater the percentage of poly-pegylated product.

In a preferred embodiment, the ChMIrp polypeptide derivative will have a single PEG moiety at the N terminus (*See* U.S. Patent No.: 8,234,784).

In another embodiment, ChMIrp polypeptides may be chemically coupled to biotin, and the biotin/Cdk11 polypeptide molecules which are conjugated are then allowed to bind to avidin, resulting in tetravalent avidin/biotin/Cdk11 polypeptide molecules. ChMIrp polypeptides may also be covalently coupled to dinitrophenol (DNP) or trinitrophenol (TNP) and the resulting conjugates precipitated with anti-DNP or anti-TNP-IgM to form decameric conjugates with a valency of 10.

Generally, conditions which may be alleviated or modulated by the administration of the present ChMIrp polypeptide derivatives include those described herein for ChMIrp polypeptides. However, the ChMIrp polypeptide derivatives disclosed herein may have additional activities, enhanced or reduced biological activity, or other characteristics, such as increased or decreased half-life, as compared to the non-derivatized molecules.

Microarray

It will be appreciated that DNA microarray technology can be utilized in accordance with the present invention. DNA microarrays are miniature, high density arrays of nucleic acids positioned on a solid support, such as glass. Each cell or element within the array has numerous copies of a single species of DNA which acts as a target for hybridization for its cognate mRNA. In expression profiling using DNA microarray technology, mRNA is first extracted from a cell or tissue sample and then converted enzymatically to fluorescently labeled cDNA. This material is hybridized to the microarray and unbound cDNA is removed by washing. The expression of discrete genes represented on the array is then visualized by quantitating the amount of labeled cDNA which is specifically bound to each target DNA. In this way, the expression of thousands of genes can be quantitated in a high throughput, parallel manner from a single sample of biological material.

This high throughput expression profiling has a broad range of applications with respect to the ChMIrp molecules of the invention, including, but not limited to: the identification and validation of ChMIrp disease-related genes as targets for therapeutics; molecular toxicology of ChMIrp molecules and inhibitors thereof; stratification of populations and generation of surrogate markers for clinical trials; and the enhancement of an Cdk11-related small molecule drug discovery by aiding in the identification of selective compounds in high throughput screens (HTS).

Selective Binding Agents

As used herein, the term "selective binding agents" refers to a molecule which has specificity for one or more ChMIrp polypeptides. Suitable

selective binding agents include, but are not limited to, antibodies and derivatives thereof, polypeptides and small molecules. Suitable selective binding agents may be prepared using methods known in the art. An exemplary selective binding agent of the present invention is capable of binding a certain portion of the ChMIrp polypeptide thereby inhibiting the activity or function of ChMIrp polypeptide.

Selective binding agents such as antibodies and antibody fragments that bind ChMIrp polypeptides are within the scope of the present invention. The antibodies may be polyclonal inlcuding monospecific polyclonal, monoclonal (MAbs), recombinant, chimeric, humanized such as CDR-grafted, human, single chain, and/or bispecific, as well as fragments, variants or derivatives thereof. Antibody fragments include those portions of the antibody which bind to an epitope on the ChMIrp polypeptide. Examples of such fragments include Fab and F(ab') fragments generated by enzymatic cleavage of full-length antibodies. Other binding fragments include those generated by recombinant DNA techniques, such as the expression of recombinant plasmids containing nucleic acid sequences encoding antibody variable regions.

Polyclonal antibodies directed toward a ChMIrp polypeptide generally are produced in animals *e.g.*, rabbits or mice, by means of multiple subcutaneous or intraperitoneal injections of ChMIrp polypeptide and an adjuvant. It may be useful to conjugate a ChMIrp polypeptide to a carrier protein that is immunogenic in the species to be immunized, such as keyhole limpet heocyanin, serum, albumin, bovine thyroglobulin, or soybean trypsin inhibitor. Also, aggregating agents such as alum are used to enhance the immune response. After immunization, the animals are bled and the serum is assayed for anti-ChMIrp polypeptide antibody titer.

Monoclonal antibodies directed toward ChMIrp are produced using any method which provides for the production of antibody molecules by continuous cell lines in culture. Examples of suitable methods for preparing monoclonal antibodies include the hybridoma methods of Kohler *et al.* (*Nature,* 256: 495-497, 1975) and the human B-cell hybridoma method (Kozbor, *J. Immunol.*, 133: 3001, 1984; Brodeur *et al.*, *Monoclonal Antibody Production Techniques and Applications*, pp. 51-63, Marcel Dekker, Inc., New York, NY, 1987).

Also provided by the invention are hybridoma cell lines which produce monoclonal antibodies reactive with ChMIrp polypeptides.

Monoclonal antibodies of the invention may be modified for use as therapeutics. One embodiment is a "chimeric" antibody in which a portion of the heavy and/or light chain is identical with or homologous to a corresponding sequence in antibodies derived from a particular species or belonging to a particular antibody class or subclass, while the remainder of the chain(s) is/are identical with or homologous to corresponding sequence in antibodies derived from another species or belonging to another antibody class or subclass. Also included are fragments of such antibodies, so long as they exhibit the desired biological activity (*See* U.S. Patent No. 4,816,567 and Morrison, *et al., Proc. Natl. Acad. Sci.* 81, 6851-6855, 1985).

In another embodiment, a monoclonal antibody of the invention is a "humanized" antibody. Methods for humanizing non-human antibodies are well known in the art. (*See* U.S. Patent Nos. 5,585,089 and 5,693,762). Generally, a humanized antibody has one or more amino acid residues introduced into it from a source which is non-human. Humanization can be performed , for example using methods described in the art (Jones *et al., Nature,* 321: 522-525, 1986; Riechmann *et al., Nature*, 332: 323-327, 1988; Verhoeyen *et al., Science,* 239: 1534-1536, 1988), by substituting at least a portion of a rodent complementarity-determining region (CDRs) for the corresponding regions of a human antibody.

Also encompassed by the invention are human antibodies which bind ChMIrp polypeptides, fragments, variants and/or derivatives. Using transgenic animals (*e.g.* mice) that are capable of producing a repertoire of human antibodies in the absence of endogenous immunoglobin production such antibodies are produced by immunization with a ChMIrp antigen (*i.e.*, having at least 6 contiguous amino acids), optionally conjugated to a carrier. *See* for example, Jakobovits *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 90: 2551-2555, 1993; Jakobovits *et al.*, *Nature* 362: 255-258, 1993; Bruggermann *et al.*, *Year in Immuno.*, 7: 33, 1993. In one method, such transgenic animals are produced by incapacitating the endogenous loci encoding the heavy and light immunoglobulin chains therein, and inserting loci encoding human heavy and light chain proteins into the genome thereof. Partially modified animals,

that is those having less than the full complement of modifications, are then cross-bred to obtain an animal having all of the desired immune system modifications. When administered an immunogen, these transgenic animals produce antibodies with human variable regions, including human (rather than *e.g.*, murine) amino acid sequences, including variable regions which are immunospecific for these antigens. *See* PCT application nos. PCT/US96/05928 and PCT/US93/06926. Additional methods are described in U.S. Patent No. 5,545,807, PCT application Nos. PCT/US91/245, PCT/GB89/01207, and in EP 546073B1 and EP 546073A1.

Human antibodies may also be produced by the expression of recombinant DNA in host cells or by expression in hybridoma cells as described herein. In an alternate embodiment, human antibodies can be produced from phage-display libraries (Hoogenboom *et al.*, *J. Mol. Biol.* 227: 381, 1991; Marks *et al.*, *J. Mol. Biol.* 222: 581, 1991). These processes mimic immune selection through the display of antibody repertoires on the surface of filamentous bacteriophage, and subsequent selection of phage by their binding to an antigen of choice. One such technique is described in PCT Application No. PCT/US98/17364, filed in the name of Adams *et al.*, which describes the isolation of high affinity and functional agonistic antibodies for MPL- and msk- receptors using such an approach.

Chimeric, CDR grafted, and humanized antibodies are typically produced by recombinant methods. Nucleic acids encoding the antibodies are introduced into host cells and expressed using materials and procedures described herein. In a preferred embodiment, the antibodies are produced in mammalian host cells, such as CHO cells. Monoclonal (e.g., human) antibodies may be produced by the expression of recombinant DNA in host cells or by expression in hybridoma cells as described herein.

For diagnostic applications, anti-ChMIrp antibodies typically will be labeled with a detectable moiety. The detectable moiety can be any one which is capable of producing, either directly or indirectly, a detectable signal. For example, the detectable moiety may be a radioisotope, such as ^{3}H, ^{14}C, ^{32}P, ^{35}S, or ^{125}I, a fluorescent or chemiluminescent compound, such as fluorescein isothiocyanate,

rhodamine, or luciferin; or an enzyme, such as alkaline phosphatase, β-galactosidase or horseradish peroxidase. (*See* Bayer *et al., Meth. Enz,* 184: 138-163, 1990).

The anti-ChMIrp antibodies of the invention may be employed in any known assay method, such as competitive binding assays, direct and indirect sandwich assays, and immunoprecipitation assays (Sola, *Monoclonal Antibodies: A Manual of Techniques*, pp. 147-158 CRC Press, Inc., 1987) for the detection and quantitation of ChMIrp polypeptides. The antibodies will bind ChMIrp polypeptides with an affinity which is appropriate for the assay method being employed.

The activity of the cell lysate or purified ChMIrp polypeptide variant is then screened in a suitable screening assay for the desired characteristic. For example, a change in the binding affinity for a ligand or immunological character of the ChMIrp polypeptide, such as affinity for a given antibody, is measured by a competitive type immunoassay. Changes in immunomodulation activity are measured by the appropriate assay. Modifications of such protein properties as redox or thermal stability hydrophobicity, susceptibility to proteolytic degradation or the tendency to aggregate with carriers or into multimers are assayed by methods well known to the ordinarily skilled artisan. Competitive binding assays rely on the ability of a labeled standard (*e.g.*, a ChMIrp polypeptide, or an immunologically reactive portion thereof) to compete with the test sample analyte (a ChMIrp polypeptide) for binding with a limited amount of anti-ChMIrp antibody. The amount of a ChMIrp polypeptide in the test sample is inversely proportional to the amount of standard that becomes bound to the antibodies. To facilitate determining the amount of standard that becomes bound, the antibodies generally are insolubilized before or after the competition, so that the standard and analyte that are bound to the antibodies may conveniently be separated from the standard and analyte which remain unbound.

Sandwich assays typically involve the use of two antibodies, each capable of binding to a different immunogenic portion, or epitope, of the protein to be detected and/or quantified. In a sandwich assay, the test sample analyte is typically bound by a first antibody which is immobilized on a solid support, and thereafter a second antibody binds to the analyte, thus forming an insoluble three part complex (*See, e.g.*, U.S. Patent No. 4,376,110). The second antibody may itself be

labeled with a detectable moiety (direct sandwich assays) or may be measured using an anti-immunoglobulin antibody that is labeled with a detectable moiety (indirect sandwich assays). For example, one type of sandwich assay is an enzyme-linked immunosorbant assay (ELISA), in which case the detectable moiety is an enzyme.

The selective binding agent, including anti-ChMIrp antibodies, are also useful for *in vivo* imaging. An antibody labeled with a detectable moiety may be administered to an animal, preferably into the bloodstream, and the presence and location of the labeled antibody in the host is assayed. The antibody may be labeled with any moiety that is detectable in an animal, whether by nuclear magnetic resonance, radiology, or other detection means known in the art.

Selective binding agents of the invention, including anti-ChMIrp antibodies, may be used as therapeutics. These therapeutic agents are generally agonists or antagonists, in that they either enhance or reduce, respectively, at least one of the biological activities of a ChMIrp polypeptide. In one embodiment, antagonist antibodies of the invention are antibodies or binding fragments thereof which are capable of specifically binding to a ChMIrp polypeptide and which are capable of inhibiting or eliminating the functional activity of a ChMIrp polypeptide *in vivo* or *in vitro*. In preferred embodiments, an antagonist antibody will inhibit the functional activity of a ChMIrp polypeptide at least about 50%, preferably at least about 80%, more preferably 90%, and most preferably 100%. Agonist and antagonist anti-ChMIrp antibodies are identified by screening assays described below.

The invention also relates to a kit comprising ChMIrp selective binding agents (such as antibodies) and other reagents useful for detecting ChMIrp levels in biological samples. Such reagents may include a secondary activity, a detectable label, blocking serum, positive and negative control samples, and detection reagents.

ChMIrp polypeptides can be used to clone ChMIrp binding partners using an "expression cloning" strategy. Radiolabeled (^{125}I) ChMIrp polypeptide or "affinity/activity-tagged" ChMIrp polypeptide (such as an Fc fusion or an alkaline phosphatase fusion) can be used in binding assays to identify a cell type or a cell line or tissue that expresses ChMIrp binding partners. RNA isolated from such cells or

tissues can then be converted to cDNA, cloned into a mammalian expression vector, and transfected into mammalian cells (for example, COS, or 293) to create an expression library. Radiolabeled or tagged ChMIrp polypeptide can then be used as an affinity reagent to identify and isolate the subset of cells in this library expressing ChMIrp binding partners. DNA is then isolated from these cells and transfected into mammalian cells to create a secondary expression library in which the fraction of cells expressing ChMIrp binding partners would be many-fold higher than in the original library. This enrichment process can be repeated iteratively until a single recombinant clone containing an ChMIrp binding partner is isolated. Isolation of ChMIrp binding partners is useful for identifying or developing novel agonists and antagonists of the ChMIrp signaling pathway. Such agonists and antagonists include ChMIrp binding partners, cyclins, Cdk inhibitors, Cdk cofactors, anti-Cdk11 binding partner antibodies, small molecules or antisense oligonucleotides.

Diagnostic Kits and Reagents

This invention also contemplates use of ChMIrp polypeptides, fragments thereof, peptides, binding compositions, and their fusion products in a variety of diagnostic kits and methods for detecting the presence of receptors and/or antibodies, or ligands. Typically the kit will have a compartment containing a ChMIrp peptide or gene segment or a reagent which recognizes one or the other, *e.g.*, binding reagents.

A kit for determining the binding affinity of a receptor or test compound to ChMIrp would typically comprise a test compound; a labeled compound, for example an antibody having known binding affinity for the protein; or a source of ligand (naturally occurring or recombinant), and a means for separating bound from free labeled compound, such as a solid phase for immobilizing the ligand or receptor. Once compounds are screened, those having suitable binding affinity to the ligand or receptor can be evaluated in suitable biological assays, as are well known in the art, to determine whether they act as agonists or antagonists to the receptor. The availability of recombinant receptor polypeptides also provide well defined standards for calibrating such assays or as positive control samples.

A preferred kit for determining the concentration of, for example, ChMIrp in a sample would typically comprise a labeled compound, *e.g.*, antibody, having known binding affinity for the target, a source of ligand or receptor (naturally occurring or recombinant), and a means for separating the bound from free labeled compound, for example, a solid phase for immobilizing the ligand or receptor. Compartments containing reagents, and instructions for use or disposal, will normally be provided.

Antibodies, including antigen binding fragments, specific for the ligand or receptor, or fragments are useful in diagnostic applications to detect the presence of elevated levels of ligand, receptor, and/or its fragments. Such diagnostic assays can employ lysates, live cells, fixed cells, immunofluorescence, cell cultures, body fluids, and further can involve the detection of antigens related to the ligand or receptor in serum, or the like. Diagnostic assays may be homogeneous (without a separation step between free reagent and antigen complex) or heterogeneous (with a separation step). Various commercial assays exist, such as radioimmunoassay (RIA), enzyme-linked immunosorbent assay (ELISA), enzyme immunoassay (EIA), enzyme-multiplied immunoassay technique (EMIT), substrate-labeled fluorescent immunoassay (SLFIA), and the like. For example, unlabeled antibodies can be employed by using a second antibody which is labeled and which recognizes the primary antibody to a ligand or receptor or to a particular fragment thereof. Similar assays have also been extensively discussed in the literature (*See*, *e.g.*, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1988).

Anti-idiotypic antibodies may have similar uses to diagnose presence of antibodies against a ligand or receptor, as such may be diagnostic of various abnormal states. For example, overproduction of a ligand or receptor may result in production of various immunological reactions which may be diagnostic of abnormal physiological states, particularly in various inflammatory or allergic conditions.

Frequently, the reagents for diagnostic assays are supplied in kits, so as to optimize the sensitivity of the assay. For the subject invention, depending upon the nature of the assay, the protocol, and the label, either labeled or unlabeled

antibody or labeled ligand or receptor is provided. This is usually in conjunction with other additives, such as buffers, stabilizers, materials necessary for signal production such as substrates for enzymes, and the like. Preferably, the kit will also contain instructions for proper use and disposal of the contents after use. Typically the kit has compartments or containers for each useful reagent. Desirably, the reagents are provided as a dry lyophilized powder, where the reagents may be reconstituted in an aqueous medium providing appropriate concentrations of reagents for performing the assay.

The aforementioned constituents of the drug screening and the diagnostic assays may be used without modification or may be modified in a variety of ways. For example, labeling may be achieved by covalently or non-covalently joining a moiety which directly or indirectly provides a detectable signal. In any of these assays, the ligand, test compound, receptor, or antibodies thereto can be labeled either directly or indirectly. Possibilities for direct labeling include label groups: radiolabels such as ^{125}I, enzymes (U.S. Pat. No. 3,645,090) such as peroxidase and alkaline phosphatase, and fluorescent labels (U.S. Pat. No. 3,940,475) capable of monitoring the change in fluorescence intensity, wavelength shift, or fluorescence polarization. Possibilities for indirect labeling include biotinylation of one constituent followed by binding to avidin coupled to one of the above label groups.

There are also numerous methods of separating bound from the free ligand, or alternatively bound from free test compound. The ligand or receptor can be immobilized on various matrixes, perhaps with detergents or associated lipids, followed by washing. Suitable matrixes include plastic such as an ELISA plate, filters, and beads. Methods of immobilizing the ligand or receptor to a matrix include, without limitation, direct adhesion to plastic, use of a capture antibody, chemical coupling, and biotin-avidin. The last step in this approach may involve the precipitation of antigen/antibody complex by any of several methods including those utilizing, *e.g.*, an organic solvent such as polyethylene glycol or a salt such as ammonium sulfate. Other suitable separation techniques include, without limitation, the fluorescein antibody magnetizable particle method described in Rattle *et al. (Clin.*

Chem., 30: 1457-1461, 1984), and the double antibody magnetic particle separation as described in U.S. Pat. No. 4,659,6178, incorporated herein by reference.

Methods for linking proteins or their fragments to the various labels have been extensively reported in the literature and do not require detailed discussion here. Many of the techniques involve the use of activated carboxyl groups either through the use of carbodiimide or active esters to form peptide bonds, the formation of thioethers by reaction of a mercapto group with an activated halogen such as chloroacetyl, or an activated olefin such as maleimide, for linkage, or the like. Fusion proteins will also find use in these applications.

Nucleic acid molecules of the invention may be used to map the locations of the ChMIrp gene and related genes on chromosomes. Mapping may be done by techniques known in the art, such as PCR amplification, *in situ* hybridization, and FISH.

This invention is also related to the use of all or part of the ChMIrp gene as part of a diagnostic assay for detecting diseases or susceptibility to diseases related to the presence of mutated ChMIrp gene.

Individuals carrying mutations in the human ChMIrp gene may be detected at the DNA level by a variety of techniques. Nucleic acids for diagnosis may be obtained from a patient's cells, such as from blood, urine, saliva, tissue biopsy and autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR (Saiki *et al.*, *Nature*, 324: 163-166, 1986), prior to analysis. RNA or cDNA may also be used for the same purpose. As an example, PCR primers complementary to the nucleic acid encoding ChMIrp polypeptide can be used to identify and analyze ChMIrp mutations. For example, deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to radiolabeled ChMIrp RNA or alternatively radiolabeled ChMIrp antisense DNA sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase A digestion or by differences in melting temperatures.

Genetic testing based on DNA sequence differences may be achieved by detection of alteration in electrophoretic mobility of DNA fragments in gels with or without denaturing agents. Small sequence deletions and insertions can be visualized by high resolution gel electrophoresis. DNA fragments of different sequences may be distinguished on denaturing, formamide gradient gels in which the mobilities of different DNA fragments are retarded in the gel at different positions according to their specific melting or partial melting temperatures (*See*, *e.g.*, Myers *et al., Science*, 230: 1242, 1985).

Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (*e.g.*, Cotton *et al., Proc. Natl. Acad. Sci., USA*, 85: 4397-4401, 1985).

Thus, the detection of a specific DNA sequence may be achieved by methods such as hybridization, RNase protection, chemical cleavage, direct DNA sequencing or the use of restriction enzymes, (*e.g.*, Restriction Fragment Length Polymorphisms (RFLP)) and Southern blotting of genomic DNA.

In addition to more conventional gel-electrophoresis and DNA sequencing, mutations can also be detected by *in situ* analysis.

The present invention also relates to a diagnostic assay for detecting altered levels of ChMIrp polypeptide in various tissues since an over-expression of the proteins compared to normal control tissue samples may detect the presence of a disease or susceptibility to a disease. Assays used to detect levels of ChMIrp polypeptide in a sample derived from a host are well-known to those of skill in the art and include radioimmunoassays, competitive-binding assays, Western Blot analysis, ELISA assays and "sandwich" assay. An ELISA assay (Coligan, *et al., Current Protocols in Immunology*, 1(2), Chapter 6, 1991) partially comprises preparing an antibody specific to the ChMIrp antigen, preferably a monoclonal antibody. In addition, a reporter antibody is prepared against the monoclonal antibody. The reporter antibody is attached to a detectable reagent such as radioactivity, fluorescence or in this example a horseradish peroxidase enzyme. A sample is now removed from a host and incubated on a solid support, *e.g.*, a polystyrene dish, that binds the proteins in the sample. Any free protein binding sites on the dish are then

covered by incubating with a non-specific protein like bovine serum albumin (BSA). Next, the monoclonal antibody is incubated in the dish during which time the monoclonal antibodies attach to any ChMIrp polypeptides attached to the polystyrene dish. All unbound monoclonal antibody is washed out with buffer. The reporter antibody linked to horseradish peroxidase is now placed in the dish resulting in binding of the reporter antibody to any monoclonal antibody bound to ChMIrp. Unattached reporter antibody is then washed out. Peroxidase substrates are then added to the dish and the amount of color developed in a given time period is a measurement of the amount of ChMIrp polypeptide present in a given volume of patient sample when compared against a standard curve.

A competition assay may be employed wherein antibodies specific to ChMIrp are attached to a solid support and labeled ChMIrp and a sample derived from the host are passed over the solid support and the amount of label detected, for example, by liquid scintillation chromotagraphy, can be correlated to a quantity of ChMIrp in the sample. In addition, a "sandwich" immunoassay as described above may also be carried out to quantify the amount of ChMIrp ligand in a biological sample.

The sequences of the present invention are also valuable for chromosome identification and mapping. The sequence can be specifically targeted to and can hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome wherein a gene can be localized. Few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking chromosomal location. The mapping of DNAs to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the cDNA. Computer analysis of the 3'-untranslated region of the sequence is used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers are then used for PCR screening of somatic cell hybrids

containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the primer will yield an amplified fragment.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular DNA to a particular chromosome. Using the present invention with the same oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes or pools of large genomic clones in an analogous manner. Other mapping strategies that can similarly be used to map ChMIrp ligand to its chromosome include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes and preselection by hybridization to construct chromosome specific-cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a cDNA clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with cDNA as short as 500 or 600 bases; however, clones larger than 2,000 bp have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. FISH requires use of genomic clones or clones from which the express sequence tag (EST) was derived, and the longer the better. For example, 2,000 bp is good, 4,000 is better, and more than 4,000 is probably not necessary to get good results a reasonable percentage of the time. For a review of this technique *see* Verma *et al.* (*Human Chromosomes: A Manual of Basic Techniques*, Pergamon Press, New York, NY, 1988).

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in V. McKusick, *Mendelian Inheritance in Man* (available on line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

Next, it is necessary to determine the differences in the cDNA or genomic sequence between affected and unaffected individuals. If a mutation is

observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

With current resolution of physical mapping and genetic mapping techniques, a cDNA precisely localized to a chromosomal region associated with the disease could be one of between 50 and 500 potential causative genes. (This assumes 1 megabase mapping resolution and one gene per 20 kb).

The nucleic acid molecule(s) of the present invention are also useful as antisense inhibitors of ChMIrp expression. Such inhibition may be effected by nucleic acid molecules which are complementary to and hybridize to expression control sequences (triple helix formation) or to ChMIrp mRNA. Antisense probes may be designed by available techniques using the sequence of ChMIrp disclosed herein. Antisense inhibitors provide information relating to the decrease or absence of a ChMIrp polypeptide in a cell or organism.

The nucleic acid molecules of the invention may be used for gene therapy. Nucleic acid molecules which express ChMIrp *in vivo* provide information relating to the effects of the polypeptide in cells or organisms. ChMIrp nucleic acid molecules, fragments, and/or derivatives that do not themselves encode biologically active polypeptides may be useful as hybridization probes in diagnostic assays to test, either qualitatively or quantitatively, for the presence of ChMIrp DNA or corresponding RNA in mammalian tissue or bodily fluid samples.

ChMIrp polypeptide fragments, variants, and/or derivatives, whether biologically active or not, are useful for preparing antibodies that bind to a ChMIrp polypeptide. The antibodies may be used for *in vivo* and *in vitro* diagnostic purposes, such as in labeled form to detect the presence of ChMIrp polypeptide in a body fluid or cell sample. The antibodies may bind to a ChMIrp polypeptide so as to diminish or block at least one activity characteristic of a ChMIrp polypeptide, or may bind to a polypeptide to increase an activity.

Genetically Engineered Non-Human Animals

Additionally included within the scope of the present invention non-human animals such as mice, rats, or other rodents, rabbits, goats, or sheep or

other farm animals, in which the gene (or genes) encoding ChMIrp polypeptides in which either the native form of the gene(s) for that mammal or a heterologous ChMIrp polypeptide gene(s) is (are) over expressed by the mammal, thereby creating a "transgenic" mammal. Such transgenic mammals may be prepared using well known methods such as those described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122.

Additionally included within the scope of the present invention are non-human animals such as mice, rats, or other rodents, rabbits, goats, or sheep or other farm animals, in which the gene (or genes) encoding a native ChMIrp polypeptide has (have) been disrupted ("knocked out") such that the level of expression of this gene or genes is (are) significantly decreased or completely abolished. Such mammals may be prepared using techniques and methods such as those described in U.S. Patent No. 5,557,032, incorporated herein by reference.

The present invention further includes non-human animals in which the promoter for one or more of the ChMIrp polypeptides of the present invention is either activated or inactivated (using homologous recombination methods as described below) to alter the level of expression of one or more of the native ChMIrp polypeptides.

These non-human animalss may be used for drug candidate screening. The impact of a drug candidate on the mammal may be measured. For example, drug candidates may decrease or increase expression of the ChMIrp polypeptide gene. In certain embodiments, the amount of ChMIrp polypeptide or a fragment(s) that is produced may be measured after exposure of the mammal to the drug candidate. In certain embodiments, one may detect the actual impact of the drug candidate on the mammal. For example, over expression of a particular gene may result in, or be associated with, a disease or pathological condition. In such cases, one may test a drug candidate's ability to decrease expression of the gene or its ability to prevent or inhibit a pathological condition. In other examples, production of a particular metabolic product such as a fragment of a polypeptide, may result in, or be associated with, a disease or pathological condition. In such cases, one may test a drug

candidate's ability to decrease production of such a metabolic product or its ability to prevent or inhibit a pathological condition.

Internalizing Proteins

The *TAT* protein sequence (from HIV) can be used to internalize proteins into a cell by targeting the lipid bi-layer component of the cell membrane. *See e.g.*, Falwell *et al.*, *Proc. Natl. Acad. Sci.*, *91*: 664-668, 1994. For example, an 11 amino acid sequence (YGRKKRRQRRR; SEQ ID NO: 16) of the HIV TAT protein (termed the "protein transduction domain", or TAT PDT) has been shown to mediate delivery of large bioactive proteins such as β-galactosidase and p27Kip across the cytoplasmic membrane and the nuclear membrane of a cell. *See* Schwarze *et al.*, *Science*, *285*: 1569-1572, 1999; and Nagahara *et al.*, *Nature Medicine, 4*: 1449-1452, 1998. Schwartze *et al.* (Science, *285*: 1569-72, 1999) demonstrated that cultured cells acquired β-gal activity when exposed to a fusion of the TAT PDT and b-galactosidase. Injection of mice with the TAT-β-gal fusion proteins resulted in β-gal expression in a number of tissues, including liver, kidney, lung, heart, and brain tissue.

It will thus be appreciated that the TAT protein sequence may be used to internalize a desired protein or polypeptide into a cell. In the context of the present invention, the TAT protein sequence can be fused to another molecule such as a ChMIrp antagonist (*i.e.*: anti-ChMIrp selective binding agent or small molecule) and administered intracellularly to inhibit the activity of the ChMIrp molecule. Where desired, the ChMIrp protein itself, or a peptide fragment or modified form of ChMIrp, may be fused to such a protein transducer for administrating to cells using the procedures, described above.

Assaying for other Modulators of ChMIrp Polypeptide Activity:

In some situations, it may be desirable to identify molecules that are modulators, *i.e.*, agonists or antagonists, of the activity of ChMIrp polypeptide. Natural or synthetic molecules that modulate ChMIrp can be identified using one or more of the screening assays described below. Such molecules may be administered

either in an *ex vivo* manner, or in an *in vivo* manner by local or intravenous (iv) injection, or by oral delivery, implantation device, or the like. "Test molecule(s)" refers to the molecule(s) that is/are under evaluation for the ability to bind to a ChMIrp polypeptide and thereby modulate its activity. A test molecule will bind to a ChMIrp polypeptide with an affinity constant of at least about 10^{-6} M, peferably about 10^{-8}M, more preferably about 10^{-9}M, and even more preferably about 10^{-10}M.

Methods for identifying compounds which interact with ChMIrp polypeptides are encompassed by the invention. In general, a ChMIrp polypeptide is incubated with a test molecule under conditions which permit binding of the test molecule to ChMIrp polypeptide, and the extent of binding is measured. The test molecules may be screened in a substantially purified form or in a crude mixture. Test molecules may be nucleic acid molecules, proteins, peptides, carbohydrates, lipids or small molecular weight organic or inorganic compounds. Once a set of test molecules has been identified as binding to a ChMIrp polypeptide, the molecules may be further evaluated for their ability to increase or decrease ChMIrp activity.

Measurement of the interaction of test molecules with ChMIrp polypeptides may be carried out in several formats, including cell-based binding assays, membrane binding assays, solution-phase assays and immunoassays. In general, test molecules are incubated with a ChMIrp polypeptide for a specified period of time and the extent of binding to a ChMIrp polypetide is determined by filtration, electrochemiluminescent (ECL, ORIGEN system by IGEN), cell-based or immunoassays.

Homogeneous assay technologies for radioactivity (SPA; Amersham) and time resolved fluorescence (HTRF, Packard) can also be implemented. Binding can be detected by labeling with radioactive isotopes (^{125}I, ^{35}S, ^{3}H), fluorescent dyes (fluorescein), lanthanides such as Europeum (Eu^{3+}) chelates or cryptates, orbipyridyl-ruthenium (Ru^{2+}) complexes. It is understood that the choice of a labeled probe will depend upon the detection system used. Alternatively, a ChMIrp polypeptide may be modified with an unlabeled epitope tag (*e.g.*, biotin, peptides, His6, myc, Fc) and bound to proteins such as streptavidin, anti-peptide or anti-protein antibodies which have a detectable label as described above.

The interaction of test molecules with ChMIrp polypeptides may also be assayed directly using polyclonal or monoclonal antibodies in an immunoassay. Alternatively, modified forms of ChMIrp polypeptides containing epitope tags as described above may be used in solution and immunoassays.

In one embodiment, a ChMIrp agonist or antagonist may be a protein, peptide, carbohydrate, lipid or small molecular weight molecule which interacts with ChMIrp to regulate its activity. Potential protein antagonists of ChMIrp include antibodies which bind to active regions of the polypeptide and inhibit or eliminate at least one activity of ChMIrp. Molecules which regulate ChMIrp polypeptide expression may include nucleic acids which are complementary to nucleic acids encoding a ChMIrp polypeptide, or are complementary to nucleic acids sequences which direct or control expression of polypeptide, and which act as antisense regulators of expression.

In the event that ChMIrp polypeptides display biological activity through interaction with a binding partner (*e.g.*, a receptor or a ligand), a variety of assays may be used to measure binding of a ChMIrp polypeptide to a corresponding binding partner. These assays may be used to screen test molecules for their ability to increase or decrease the rate and/or the extent of binding of a ChMIrp polypeptide to its binding partner. In one assay, a ChMIrp polypeptide is immobilized by attachment to the bottom of the wells of a microtiter plate. Radiolabeled ChMIrp binding partner (for example, iodinated ChMIrp binding partner) and the test molecule(s) can then be added either one at a time (in either order) or simultaneously to the wells. After incubation, the wells can be washed and counted using a scintillation counter for radioactivity to determine the extent of binding to ChMIrp polypeptide by its binding partner. Typically, the molecules will be tested over a range of concentrations, and a series of control wells lacking one or more elements of the test assays can be used for accuracy in evaluation of the results. An alternative to this method involves reversing the "positions" of the proteins, *i.e.*, immobilizing ChMIrp binding partner to the microtiter plate wells, incubating with the test molecule and radiolabeled ChMIrp and determining the extent of ChMIrp binding

(*See*, *e.g.*, Chapter 18 of Ausubel *et al.*, eds., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY, 1995).

As an alternative to radio-labeling, an ChMIrp polypeptide or its binding partner may be conjugated to biotin and the presence of biotinylated protein can then be detected using streptavidin linked to an enzyme, such as horse radish peroxidase (HRP) or alkaline phosphatase (AP), that can be detected colorometrically, or by fluorescent tagging of streptavidin. An antibody directed to an ChMIrp polypeptide or to an ChMIrp binding partner and is conjugated to biotin may also be used and can be detected after incubation with enzyme-linked streptavidin linked to AP or HRP

A ChMIrp polypeptide and a ChMIrp binding partner may also be immobilized by attachment to agarose beads, acrylic beads or other types of such inert substrates. The substrate-protein complex can be placed in a solution containing the complementary protein and the test compound; after incubation, the beads can be precipitated by centrifugation, and the amount of binding between a ChMIrp polypeptide and its binding partner can be assessed using the methods described above. Alternatively, the substrate-protein complex can be immobilized in a column and the test molecule and complementary protein passed over the column. Formation of a complex between a ChMIrp polypeptide and its binding partner can then be assessed using any of the techniques set forth above, *i.e.*, radiolabeling, antibody binding, or the like.

Another *in vitro* assay that is useful for identifying a test molecule which increases or decreases formation of a complex between a ChMIrp binding protein and a ChMIrp binding partner is a surface plasmon resonance detector system such as the Biacore assay system (Pharmacia, Piscataway, NJ). The Biacore system may be carried out using the manufacturer's protocol. This assay essentially involves covalent binding of either ChMIrp or a ChMIrp binding partner to a dextran-coated sensor chip which is located in a detector. The test compound and the other complementary protein can then be injected into the chamber containing the sensor chip either simultaneously or sequentially and the amount of complementary protein that bind to each other can be assessed based on the change in molecular mass which

is physically associated with the dextran-coated side of the sensor chip; the change in molecular mass can be measured by the detector system.

In some cases, it may be desirable to evaluate two or more test compounds together for their ability to increase or decrease formation of a complex between a ChMIrp polypeptide and a ChMIrp binding partner complex. In these cases, the assays set forth above can be readily modified by adding such additional test compound(s) either simultaneous with, or subsequent to, the first test compound. The remainder of steps in the assay are as set forth above.

In vitro assays such as those described above may be used advantageously to screen rapidly large numbers of compounds for effects on complex formation by ChMIrp and ChMIrp binding partner. The assays may be automated to screen compounds generated in phage display, synthetic peptide and chemical synthesis libraries.

Compounds which increase or decrease formation of a complex between a ChMIrp polypeptide and a ChMIrp binding partner may also be screened in cell culture using cells and cell lines expressing either ChMIrp or ChMIrp binding partner. Cells and cell lines may be obtained from any mammal, but preferably will be from human or other primate, canine, or rodent sources. The binding of a ChMIrp polypeptide to cells expressing ChMIrp binding partner at the surface is evaluated in the presence or absence of test molecules and the extent of binding may be determined by, for example, flow cytometry using a biotinylated antibody to a ChMIrp binding partner. Cell culture assays may be used advantageously to further evaluate compounds that score positive in protein binding assays described above.

Cell cultures can also be used to screen the impact of a drug candidate. For example, drug candidates may decrease or increase the expression of the ChMIrp polypeptide gene. In certain embodiments, the amount of ChMIrp polypeptide or a fragment(s) that is produced may be measured after exposure of the cell culture to the drug candidate. In certain embodiments, one may detect the actual impact of the drug candidate on the cell culture. For example, the overexpression of a particular gene may have a particular impact on the cell culture. In such cases, one may test a drug candidate's ability to increase or decrease the expression of the gene or its ability to

prevent or inhibit a particular impact on the cell culture. In other examples, the production of a particular metabolic product such as a fragment of a polypeptide, may result in, or be associated with, a disease or pathological condition. In such cases, one may test a drug candidate's ability to decrease the production of such a metabolic product in a cell culture.

A yeast two hybrid system (Chien *et al.*, *Proc. Natl. Acad. Sci. USA*, 88: 9578-9583, 1991) can be used to identify novel polypeptides that bind to, or interact with, ChMIrp polypeptides. As an example, a yeast-two hybrid bait construct can be generated in a vector (such as the pAS2-1 from Clontech) which encodes a yeast GAL4-DNA binding domain fused to the ChMIrp polynucleotide. This bait construct may be used to screen human cDNA libraries wherein the cDNA library sequences are fused to GAL4 activation domains. Positive interactions will result in the activation of a reporter gene such as β-Gal. Positive clones emerging from the screening may be characterized further to identify interacting proteins.

ChMIrp Polypeptide Compositions and Administration

Chondromodulin family members are known to induce cartilage formation and bone growth. In addition, these proteins are suggested to be inhibitors of vascularization. Angiogenesis mediates many pathological conditions. These described biological activities suggest possible therapeutic uses for administration of chondromodulins.

Pharmaceutical compositions of ChMIrp polypeptides are within the scope of the present invention for prophylactic and therapeutic treatment of humans and animals for indications resulting from decreased levels of ChMIrp or where it is determined that administration of ChMIrp polypeptide will result in the amelioration or cure of the indications. Such compositions may comprise a therapeutically effective amount of a ChMIrp polypeptide and/or its binding partner, or therapeutically active fragment(s), variant(s), or derivative(s) thereof in a mixture with a pharmaceutically acceptable additives and/or carriers. Suitable formulation materials or pharmaceutically acceptable agents include, but are not limited to, antioxidants, preservatives, colors, flavoring, diluting agents, emulsifying agents,

suspending agents, solvents, fillers, bulking agents, buffers, delivery vehicles, diluents, excipients, and/or pharmaceutical adjuvants. Typically, a therapeutic compound containing ChMIrp polypeptide(s) will be administered in the form of a composition comprising purified polypeptide, fragment(s), variant(s), or derivative(s) in conjunction with one or more physiologically acceptable carriers, excipients, or diluents. For example, a suitable vehicle may be water for injection, physiological solution, or artificial cerebrospinal fluid possibly supplemented with other materials common in compositions for parenteral delivery.

Neutral buffered saline or saline mixed with serum albumin are exemplary appropriate carriers. Preferably, the product is formulated as a lyophilizate using appropriate excipients (*e.g.*, sucrose). Other standard carriers, diluents, and excipients may be included as desired. Other exemplary compositions comprise Tris buffer of about pH 7.0-8.5, or acetate buffer of about pH 4.0-5.5, which may further include sorbitol or a suitable substitute therefor. The pH of the solution should also be selected based on the relative solubility of ChMIrp ligand at various pHs.

The primary solvent in a composition may be either aqueous or non-aqueous in nature. In addition, the vehicle may contain other formulation materials for modifying or maintaining the pH, osmolarity, viscosity, clarity, color, sterility, stability, rate of dissolution, or odor of the formulation. Similarly, the composition may contain additional formulation materials for modifying or maintaining the rate of release of ChMIrp polypeptide, or for promoting the absorption or penetration of ChMIrp polypeptide.

Compositions comprising the ChMIrp polypeptide compositions can be administered parentally. Alternatively, the compositions may be administered intravenously or subcutaneously. When systemically administered, the therapeutic compositions for use in this invention may be in the form of a pyrogen-free, parentally acceptable aqueous solution. The preparation of such pharmaceutically acceptable protein solutions, with due regard to pH, isotonicity, stability and the like, is within the skill of the art.

Therapeutic formulations of ChMIrp polypeptide compositions useful for practicing the present invention may be prepared for storage by mixing the selected composition having the desired degree of purity with optional physiologically acceptable carriers, excipients, or stabilizers (*Remington's Pharmaceutical Sciences*, 18th Edition, A.R. Gennaro, *ed.*, Mack Publishing Company, 1990) in the form of a lyophilized cake or an aqueous solution.

Acceptable carriers, excipients or stabilizers are nontoxic to recipients and are preferably inert at the dosages and concentrations employed, and include buffers such as phosphate, citrate, or other organic acids; antioxidants such as ascorbic acid; low molecular weight polypeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, arginine or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; salt-forming counterions such as sodium; and/or nonionic surfactants such as Tween, pluronics or polyethylene glycol (PEG).

An effective amount of the ChMIrp polypeptide(s) composition to be employed therapeutically will depend, for example, upon the therapeutic objectives such as the indication for which the composition is being used, the route of administration (*e.g.*, whether it is administered locally or systemically), and the condition of the patient (*e.g.*, patient's general health, anaureuesis, age, weight, sex). It is essential, when determining the therapeutically effective dose, to take into account the quantity of ChMIrp or other members of the chondromodulin family secreted which are responsible for the disease as well as the quantity of endogenous ChMIrp Accordingly, it will be necessary for the therapist to titer the dosage and/or *in vivo* modify the route of administration as required to obtain the optimal therapeutic effect. A typical daily dosage may range from about 0.1 mg/kg to up to 100 mg/kg or more, depending on the factors mentioned above. Typically, a clinician will administer the composition until a dosage is reached that achieves the desired effect. The composition may therefore be administered as a single dose, or as

two or more doses which may or may not contain the same amount of ChMIrp polypeptide over time, or as a continuous infusion via implantation device or catheter.

The frequency of dosing will depend upon the pharmacokinetic parameters of the ChMIrp molecule in the formulation used. Typically, a clinician will administer the composition until a dosage is reached that achieves the desired effect. The composition may therefore be administered as a single dose, or as two or more doses (which may or may not contain the same amount of the desired molecule) over time, or as a continuous infusion via implantation device or catheter. Further refinement of the appropriate dosage is routinely made by those of ordinary skill in the art and is within the ambit of tasks routinely performed by them. Appropriate dosages may be ascertained through use of appropriate dose-response data.

As further studies are conducted, information will emerge regarding appropriate dosage levels for treatment of various conditions in various patients, and the ordinary skilled worker, considering the therapeutic context, the type of disorder under treatment, the age and general health of the recipient, will be able to ascertain proper dosing.

The ChMIrp polypeptide composition to be used for *in vivo* administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes. Where the composition is lyophilized, sterilization using these methods may be conducted either prior to or following lyophilization and reconstitution. The composition for parenteral administration ordinarily will be stored in lyophilized form or in solution.

Therapeutic compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Effective administration forms, such as (1) slow-release formulations, (2) inhalant mists, or (3) orally active formulations are also envisioned. Pharmaceutical compositions comprising thereapeutically effective dose of the ChMIrp polypeptide also may be formulated for parenteral administration. Such parenterally administered therapeutic compositions are typically in the form of a pyrogen-free, parenterally acceptable aqueous solution comprising ChMIrp in a

pharmaceutically acceptable vehicle. The ChMIrp pharmaceutical compositions also may include particulate preparations of polymeric compounds such as polylactic acid, polyglycolic acid, etc. or the introduction of ChMIrp into liposomes. Hyaluronic acid may also be used, and this may have the effect of promoting sustained duration in the circulation.

A particularly suitable vehicle for parenteral injection is sterile distilled water in which ChMIrp is formulated as a sterile, isotonic solution, properly preserved. Yet another preparation may involve the formulation of ChMIrp with an agent, such as injectable microspheres, bio-erodible particles or beads, or liposomes, that provides for the controlled or sustained release of the protein product which may then be delivered as a depot injection. Other suitable means for the introduction of ChMIrp include implantable drug delivery devices which contain the ChMIrp and/or its binding partner.

The preparations of the present invention may include other components, for example parenterally acceptable preservatives, tonicity agents, cosolvents, wetting agents, complexing agents, buffering agents, antimicrobials, antioxidants and surfactants, as are well known in the art. For example, suitable tonicity enhancing agents include alkali metal halides (preferably sodium or potassium chloride), mannitol, sorbitol and the like. Suitable preservatives include, but are not limited to, benzalkonium chloride, thimerosal, phenethyl alcohol, methylparaben, propylparaben, chlorhexidine, sorbic acid and the like. Hydrogen peroxide may also be used as preservative. Suitable cosolvents are for example glycerin, propylene glycol and polyethylene glycol. Suitable complexing agents are for example caffeine, polyvinylpyrrolidone, beta-cyclodextrin or hydroxypropyl-beta-cyclodextrin. Suitable surfactants or wetting agents include sorbitan esters, polysorbates such as polysorbate 80, tromethamine, lecithin, cholesterol, tyloxapal and the like. The buffers can be conventional buffers such as borate, citrate, phosphate, bicarbonate, or Tris-HCl.

The formulation components are present in concentration that are acceptable to the site of administration. For example, buffers are used to maintain

the composition at physiological pH or at slightly lower pH, typically within a pH range of from about 5 to about 8.

When parenteral administration is contemplated, the therapeutic compositions for use in this invention may be in the form of a pyrogen-free, parenterally acceptable aqueous solution comprising the desired ChMIrp molecule in a pharmaceutically acceptable vehicle. A particularly suitable vehicle for parenteral injection is sterile distilled water in which a ChMIrp molecule is formulated as a sterile, isotonic solution, properly preserved. Yet another preparation can involve the formulation of the desired molecule with an agent, such as injectable microspheres, bio-erodible particles, polymeric compounds (polylactic acid, polyglycolic acid), beads, or liposomes, that provides for the controlled or sustained release of the product which may then be delivered via a depot injection. Hyaluronic acid may also be used, and this may have the effect of promoting sustained duration in the circulation. Other suitable means for the introduction of the desired molecule include implantable drug delivery devices.

In one embodiment, a pharmaceutical composition may be formulated for inhalation. For example, ChMIrp molecule may be formulated as a dry powder for inhalation. ChMIrp polypeptide or ChMIrp nucleic acid molecule inhalation solutions may also be formulated with a propellant for aerosol delivery. In yet another embodiment, solutions may be nebulized. Pulmonary administration is further described in PCT Application No. PCT/US94/001875, which described pulmonary delivery or chemically modified protiens.

It is also contemplated that certain formulations containing ChMIrp may be administered orally. In one embodiment of the present invention, ChMIrp molecules which are administered in this fashion can be formulated with or without those carriers customarily used in the compounding of solid dosage forms such as tablets and capsules. For example, a capsule may be designed to release the active portion of the formulation at the point in the gastrointestinal tract when bioavailability is maximized and pre-systemic degradation is minimized. Additional agents can be included to facilitate absorption of the ChMIrp molecule. Diluents, flavorings, low

melting point waxes, vegetable oils, lubricants, suspending agents, tablet disintegrating agents, and binders may also be employed.

Another pharmaceutical composition may involve an effective quantity of ChMIrp molecules in a mixture with non-toxic excipients which are suitable for the manufacture of tablets. By dissolving the tablets in sterile water, or other appropriate vehicle, solutions can be prepared in unit dose form. Suitable excipients include, but are not limited to, inert diluents, such as calcium carbonate, sodium carbonate or bicarbonate, lactose, or calcium phosphate; or binding agents, such as starch, gelatin, or acacia; or lubricating agents such as magnesium stearate, stearic acid, or talc.

Additional ChMIrp pharmaceutical compositions will be evident to those skilled in the art, including formulations involving ChMIrp polypeptides in sustained- or controlled-delivery formulations. Techniques for formulating a variety of other sustained- or controlled-delivery means, such as liposome carriers, bio-erodible microparticles or porous beads and depot injections, are also known to those skilled in the art. *See*, for example, PCT/US93/00829 which describes controlled release of porous polymeric microparticles for the delivery of pharmaceutical compositions.

Once the pharmaceutical composition has been formulated, it may be stored in sterile vials as a solution, suspension, gel, emulsion, solid, or a dehydrated or lyophilized powder. Such formulations may be stored either in a ready-to-use form or in a form (*e.g.*, lyophilized) requiring reconstitution prior to administration.

In a specific embodiment, the present invention is directed to kits for producing a single-dose administration unit. The kits may each contain both a first container having a dried protein and a second container having an aqueous formulation. Also included within the scope of this invention are kits containing single and multi-chambered pre-filled syringes (*e.g.*, liquid syringes and lyosyringes).

Regardless of the manner of administration, the specific dose may be calculated according to body weight, body surface area or organ size. Further refinement of the calculations necessary to determine the appropriate dosage for treatment involving each of the above mentioned formulations is routinely made by those of ordinary skill in the art and is within the ambit of tasks routinely performed

by them. Appropriate dosages may be ascertained through use of appropriate dose-response data.

The route of administration of the composition is in accord with known methods, *e.g.* oral, injection or infusion by intravenous, intraperitoneal, intracerebral (intraparenchymal), intraventricular, intramuscular, intraocular, intraarterial, or intralesional routes, or by sustained release systems or implantation device which may optionally involve the use of a catheter. Where desired, the compositions may be administered continuously by infusion, bolus injection or by implantation device. Alternatively or additionally, the composition may be administered locally via implantation into the affected area of a membrane, sponge, or other appropriate material on to which ChMIrp polypeptide has been absorbed.

One may further provide the present pharmaceutical compositions by pulmonary administration, *see*, *e.g.*, International Publication No: WO 94/20069, which discloses pulmonary delivery of chemically modified proteins. For pulmonary delivery, the particle size should be suitable for delivery to the distal lung. For example, the particle size may be from 1 mm to 5 mm, however, larger particles may be used, for example, if each particle is fairly porous. Alternatively or additionally, the composition may be administered locally via implantation into the affected area of a membrane, sponge, or other appropriate material on to which receptor polypeptide has been absorbed or encapsulated. Where an implantation device is used, the device may be implanted into any suitable tissue or organ, and delivery may be directly through the device via bolus, or via continuous administration, or via catheter using continuous infusion.

ChMIrp polypeptide and/or its binding partner may also be administered in a sustained release formulation or preparation. Suitable polymer compositions preferably have intrinsic and controllable biodegradability so that they persist for about a week to about six months; are non-toxic containing no significant toxic monomers and degrading into non-toxic components; are biocompatible, are chemically compatible with substances to be delivered, and tend not to denature the active substance; are sufficiently porous to allow the incorporation of biologically active molecules and their subsequent liberation from the polymer by diffusion,

erosion or a combination thereof; are able to remain at the site of the application by adherence or by geometric factions, such as being formed in place or softened and subsequently molded or formed into microparticles which are trapped at a desired location; are capable of being delivered by techniques of minimum invasivity such as by catheter, laparoscope or endoscope. Sustained release matrices include polyesters, hydrogels, polylactides (U.S. 3,773,919; EP 58,481), copolymers of L-glutamic acid and gamma ethyl-L-glutamate (Sidman *et al, Biopolymers*, 22: 547-556, 1983), poly (2-hydroxyethyl-methacrylate) (Langer *et al., J. Biomed. Mater. Res.*, 15: 167-277,1981; and Langer, *Chem. Tech.*, 12:98-105, 1982), ethylene vinyl acetate (Langer *et al.*, *supra*) or poly-D(-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also may include liposomes, which can be prepared by any of several methods known in the art (e.g., Eppstein *et al., Proc. Natl. Acad. Sci. USA*, 82: 3688-3692, 1985; EP 36,676; EP 88,046; EP 143,949).

The ChMIrp polypeptides, variants, derivatives or fragments thereof, may be employed alone, together, or in combination with other pharmaceutical compositions. The ChMIrp polypeptides, fragments, variants, and derivatives may be used in combination with cytokines, cytokine inhibitors, growth factors, antibiotics, anti-inflammatories, and/or chemotherapeutic agents as is appropriate for the indication being treated

In some cases, it may be desirable to use ChMIrp pharmaceutical compositions in an *ex vivo* manner. In such instances, cells, tissues, or organs that have been removed from the patient are exposed to ChMIrp pharmaceutical compositions after which the cells, tissues and/or organs are subsequently implanted back into the patient.

In other cases, a ChMIrp polypeptide can be delivered by implanting certain cells that have been genetically engineered, using methods such as those described herein, to express and secrete the polypeptide. Such cells may be animal or human cells, and may autologous, heterologous, or xenogeneic. Optionally, the cells may be immortalized. In order to decrease the chance of an immunological response, it the cells may be encapsulated to avoid infiltration of surrounding tissues. The encapsulation materials are typically biocompatible, semi-permeable polymeric

enclosures or membranes that allow the release of the protein product(s) but prevent the destruction of the cells by the patient's immune system or by other detrimental factors from the surrounding tissues.

Methods used for membrane encapsulation of cells are familiar to the skilled artisan, and preparation of encapsulated cells and their implantation in patients may be accomplished without undue experimentation (*See, e.g.*, U.S. Patent Nos. 4,892,538; 5,011,472; and 5,106,627). A system for encapsulating living cells is described in International Publication No: WO 91/10425. Techniques for formulating a variety of other sustained or controlled delivery means, such as liposome carriers, bio-erodible particles or beads, are also known to those in the art, and are described, for example, in U.S. Patent No. 5,653,975. The cells, with or without encapsulation, may be implanted into suitable body tissues or organs of the patient.

As discussed above, it may be desirable to treat isolated cell populations such as stem cells, lymphocytes, red blood cells, chondrocytes, neurons, and the like; add as appropriate with one or more ChMIrp polypeptides, variants, derivatives and/or fragments. This can be accomplished by exposing the isolated cells to the polypeptide, variant, derivative, or fragment directly, where it is in a form that is permeable to the cell membrane.

Additional embodiments of the present invention relate to cells and methods (*e.g.*, homologous recombination and/or other recombinant production methods) for both the *in vitro* production of therapeutic polypeptides and for the production and delivery of therapeutic polypeptides by gene therapy or cell therapy.

Homologous Recombination

It is further envisioned that ChMIrp polypeptide may be produced by homologous recombination, or with recombinant production methods utilizing control elements introduced into cells already containing DNA encoding ChMIrp polypeptide. For example, homologous recombination methods may be used to modify a cell that contains a normally transcriptionally silent ChMIrp gene, or under expressed gene, and thereby produce a cell which expresses therapeutically

efficacious amounts of ChMIrp. Homologous recombination is a technique originally developed for targeting genes to induce or correct mutations in transcriptionally active genes (Kucherlapati *et al.*, *Prog. in Nucl. Acid Res. & Mol. Biol.*, 36: 301, 1989). The basic technique was developed as a method for introducing specific mutations into specific regions of the mammalian genome (Thomas *et al.*, *Cell*, 44: 419-428, 1986; Thomas and Capecchi, *Cell*, 51: 503-512, 1987; Doetschman *et al.*, *Proc. Natl. Acad. Sci.*, 85: 8583-8587, 1988) or to correct specific mutations within defective genes (Doetschman *et al.*, *Nature*, 330: 576-578, 1987). Exemplary homologous recombination techniques are described in U.S. Patent No. 5,272,071; EP 91 90 3051; EP Publication No. 505 500; PCT/US90/07642; International Publication No. WO 91/09955.

Through homologous recombination, the DNA sequence to be inserted into the genome can be directed to a specific region of the gene of interest by attaching it to targeting DNA. The targeting DNA is a nucleotide sequence that is complementary (homologous) to a region of the genomic DNA. Small pieces of targeting DNA that are complementary to a specific region of the genome are put in contact with the parental strand during the DNA replication process. It is a general property of DNA that has been inserted into a cell to hybridize, and therefore, recombine with other pieces of endogenous DNA through shared homologous regions. If this complementary strand is attached to an oligonucleotide that contains a mutation or a different sequence or an additional nucleotide, it too is incorporated into the newly synthesized strand as a result of the recombination. As a result of the proofreading function, it is possible for the new sequence of DNA to serve as the template. Thus, the transferred DNA is incorporated into the genome.

Attached to these pieces of targeting DNA are regions of DNA which may interact with or control the expression of a ChMIrp polypeptide, *e.g.*, flanking sequences. For example, a promoter/enhancer element, a suppressor, or an exogenous transcription modulatory element is inserted in the genome of the intended host cell in proximity and orientation sufficient to influence the transcription of DNA encoding the desired ChMIrp polypeptide. The control element controls a portion of the DNA present in the host cell genome. Thus, the expression of the desired

ChMIrp polypeptide may be achieved not by transfection of DNA that encodes the ChMIrp gene itself, but rather by the use of targeting DNA (containing regions of homology with the endogenous gene of interest) coupled with DNA regulatory segments that provide the endogenous gene sequence with recognizable signals for transcription of a ChMIrp polypeptide.

In an exemplary method, the expression of a desired targeted gene in a cell (*i.e.*, a desired endogenous cellular gene) is altered by the introduction,via homologous recombination into the cellular genome at a preselected site, by the introduction of DNA which includes at least a regulatory sequence, an exon and a splice donor site. These components are introduced into the chromosomal (genomic) DNA in such a manner that this, in effect, results in production of a new transcription unit (in which the regulatory sequence, the exon and the splice donor site present in the DNA construct are operatively linked to the endogenous gene). As a result of the introduction of these components into the chromosomal DNA, the expression of the desired endogenous gene is altered.

Altered gene expression, as used described herein, encompasses activating (or causing to be expressed) a gene which is normally silent (unexpressed) in the cell as obtained, as well as increasing the expression of a gene which is not expressed at physiologically significant levels in the cell as obtained. The embodiments further encompass changing the pattern of regulation or induction such that it is different from the pattern of regulation or induction that occurs in the cell as obtained, and reducing (including eliminating) expression of a gene which is expressed in the cell as obtained.

One method by which homologous recombination can be used to increase, or cause, ChMIrp polypeptide production from a cell's endogenous ChMIrp gene involves first using homologous recombination to place a recombination sequence from a site-specific recombination system (*e.g.*, Cre/loxP, FLP/FRT) (Sauer *et al., Current Opinion In Biotechnology, 5*:521-527, 1994; Sauer *et al., Methods In Enzymology, 225*:890-900, 1993) upstream (that is, 5' to) of the cell's endogenous genomic ChMIrp polypeptide coding region. A plasmid containing a recombination site homologous to the site that was placed just upstream of the

genomic ChMIrp polypeptide coding region is introduced into the modified cell line along with the appropriate recombinase enzyme. This recombinase causes the plasmid to integrate, via the plasmid's recombination site, into the recombination site located just upstream of the genomic ChMIrp polypeptide coding region in the cell line (Baubonis and Sauer, *Nucleic Acids Res., 21*:2025-2029, 1993; O'Gorman *et al., Science, 251*:1351-1355, 1991). Any flanking sequences known to increase transcription (*e.g.*, enhancer/promoter, intron, translational enhancer), if properly positioned in this plasmid, would integrate in such a manner as to create a new or modified transcriptional unit resulting in de novo or increased ChMIrp polypeptide production from the cell's endogenous ChMIrp gene.

A further method to use the cell line in which the site specific recombination sequence had been placed just upstream of the cell's endogenous genomic ChMIrp polypeptide coding region is to use homologous recombination to introduce a second recombination site elsewhere in the cell line's genome. The appropriate recombinase enzyme is then introduced into the two-recombination-site cell line, causing a recombination event (deletion, inversion, translocation) (Sauer *et al., Current Opinion In Biotechnology, supra*, 1994; *Sauer, Methods In Enzymology, supra*, 1993) that would create a new or modified transcriptional unit resulting in de novo or increased ChMIrp polypeptide production from the cell's endogenous ChMIrp gene.

An additional approach for increasing, or causing, the expression of ChMIrp polypeptide from a cell's endogenous ChMIrp gene involves increasing, or causing, the expression of a gene or genes (*e.g.*, transcription factors) and/or decreasing the expression of a gene or genes (*e.g.*, transcriptional repressors) in a manner which results in de novo or increased ChMIrp polypeptide production from the cell's endogenous ChMIrp gene. This method includes the introduction of a non-naturally occurring polypeptide (*e.g.*, a polypeptide comprising a site specific DNA binding domain fused to a transcriptional factor domain) into the cell such that de novo or increased ChMIrp polypeptide production from the cell's endogenous ChMIrp gene results.

The present invention further relates to DNA constructs useful in the method of altering expression of a target gene. In certain embodiments, the exemplary DNA constructs comprise: (a) one or more targeting sequences; (b) a regulatory sequence; (c) an exon; and (d) an unpaired splice-donor site. The targeting sequence in the DNA construct directs the integration of elements (a) - (d) into a target gene in a cell such that the elements (b) - (d) are operatively linked to sequences of the endogenous target gene. In another embodiment, the DNA constructs comprise: (a) one or more targeting sequences, (b) a regulatory sequence, (c) an exon, (d) a splice-donor site, (e) an intron, and (f) a splice-acceptor site, wherein the targeting sequence directs the integration of elements (a) - (f) such that the elements of (b) - (f) are operatively linked to the endogenous gene. The targeting sequence is homologous to the preselected site in the cellular chromosomal DNA with which homologous recombination is to occur. In the construct, the exon is generally 3′ of the regulatory sequence and the splice-donor site is 3′of the exon.

If the sequence of a particular gene is known, such as the nucleic acid sequence of ChMIrp polypeptide presented herein, a piece of DNA that is complementary to a selected region of the gene can be synthesized or otherwise obtained, such as by appropriate restriction of the native DNA at specific recognition sites bounding the region of interest. This piece serves as a targeting sequence upon insertion into the cell and will hybridize to its homologous region within the genome. If this hybridization occurs during DNA replication, this piece of DNA, and any additional sequence attached thereto, will act as an Okazaki fragment and will be incorporated into the newly synthesized daughter strand of DNA. The present invention, therefore, includes nucleotides encoding a ChMIrp polypeptide, which nucleotides may be used as targeting sequences.

Alternatively, gene therapy can be employed as described below.

ChMIrp Cell Therapy and Gene Therapy

ChMIrp cell therapy, *e.g.*, the implantation of cells producing ChMIrp, is also contemplated. This embodiment involves implanting into patients cells capable of synthesizing and secreting a biologically active form of ChMIrp

polypeptide. Such ChMIrp polypeptide-producing cells can be cells that are natural producers of ChMIrp or may be recombinant cells whose ability to produce ChMIrp polypeptides has been augmented by transformation with a gene encoding the desired ChMIrp polypeptides or with a gene augmenting the expression of ChMIrp polypeptide. Such a modification may be accomplished by means of a vector suitable for delivering the gene as well as promoting its expression and secretion. In order to minimize a potential immunological reaction in patients being administered a ChMIrp polypeptide, as may occur with the administration of a polypeptide of a foreign species, it is preferred that the natural cells producing ChMIrp polypeptide be of human origin and produce human ChMIrp. Likewise, it is preferred that the recombinant cells producing ChMIrp polypeptide be transformed with an expression vector containing a gene encoding a human ChMIrp polypeptide.

Implanted cells may be encapsulated to avoid infiltration of surrounding tissue. Human or non-human animal cells may be implanted in patients in biocompatible, semipermeable polymeric enclosures or membranes that allow release of ChMIrp polypeptide, but prevent the destruction of the cells by the patient's immune system or by other detrimental factors from the surrounding tissue. Alternatively, the patient's own cells, transformed to produce ChMIrp polypeptides *ex vivo*, may be implanted directly into the patient without such encapsulation.

Techniques for the encapsulation of living cells are known in the art, and the preparation of the encapsulated cells and their implantation in patients may be routinely accomplished. For example, Baetge *et al.* (International Publication No. WO 95/05452; International Application No. PCT/US94/09299) describe membrane capsules containing genetically engineered cells for the effective delivery of biologically active molecules. The capsules are biocompatible and are easily retrievable. The capsules encapsulate cells transfected with recombinant DNA molecules comprising DNA sequences coding for biologically active molecules operatively linked to promoters that are not subject to down-regulation *in vivo* upon implantation into a mammalian host. The devices provide for delivery of the molecules from living cells to specific sites within a recipient In addition, *see* U.S. Patent Numbers 4,892,538, 5,011,472, and 5,106,627. A system for encapsulating

living cells is described in International Application WO 91/10425 of Aebischer *et al.*, International Application WO 91/10470 of Aebischer *et al.*; Winn *et al.*, *Exper. Neurol.*, 113: 322-329, 1991; Aebischer *et al.*, *Exper. Neurol.*, 111: 269-275, 1991; and Tresco *et al.*, *ASAIO*, 38: 17-23, 1992.

In vivo and *in vitro* gene therapy delivery of ChMIrp is also encompassed by the present invention. *In vivo* gene therapy may be accomplished by introducing the gene encoding ChMIrp into cells via local injection of a polynucleotide molecule or other appropriate delivery vectors (Hefti, *J. Neurobiology*, *25*: 1418-1435, 1994). For example, a polynucleotide molecule encoding ChMIrp may be contained in an adeno-associated virus vector for delivery to the targeted cells (International Publication No. WO 95/34670; International Application No. PCT/US95/07178). The recombinant adeno-associated virus (AAV) genome typically contains AAV inverted terminal repeats flanking a DNA sequence encoding ChMIrp operably linked to functional promoter and polyadenylation sequences.

Alternative viral vectors include, but are not limited to, retrovirus, adenovirus, herpes simplex virus and papilloma virus vectors. U.S. Patent No. 5,672,344 describes an *in vivo* viral-mediated gene transfer system involving a recombinant neurotrophic HSV-1 vector. U.S. Patent No. 5,399,346, provides examples of a process for providing a patient with a therapeutic protein by the delivery of human cells which have been treated *in vitro* to insert a DNA segment encoding a therapeutic protein. Additional materials and methods for the practice of gene therapy techniques are described in U.S. Patent No. 5,631,236; gene therapy involving adenoviral vectors are described in U.S. Patent No 5,672,510; and gene therapy involving the use of retroviral vectors are described in U.S. Patent No. 5,635,399.

Nonviral delivery methods include liposome-mediated transfer, naked DNA delivery (direct injection), receptor-mediated transfer (ligand-DNA complex), electroporation, calcium phosphate precipitation and microparticle bombardment (*e.g.*, gene gun). Gene therapy materials and methods may also include inducible promoters, tissue-specific enhancer-promoters, DNA sequences designed for

site-specific integration, DNA sequences capable of providing a selective advantage over the parent cell, labels to identify transformed cells, negative selection systems and expression control systems (safety measures), cell-specific binding agents (for cell targeting), cell-specific internalization factors, transcription factors to enhance expression by a vector as well as methods of vector manufacture. Such additional methods and materials for the practice of gene therapy techniques are described in U.S. Patent No. 4,970,154, International Application No. WO 9640958; U.S. Patent No. 5,679,559; U.S. 5,676,954; U.S. Patent No. 5,593,875; and U.S. Patent No. 4,945,050. Expression control techniques include chemical induced regulation (*e.g.*, International Application Nos. WO 9641865 and WO 9731899), the use of a progesterone antagonist in a modified steroid hormone receptor system (*e.g.*, U.S. Patent No. 5,364,791), ecdysone control systems (*e.g.*, International Application No. WO 9637609), and positive tetracycline-controllable transactivators (*e.g.*, U.S. Patent Nos. 5,589,362; 5,650,298; and 5,654,168).

In vivo and *in vitro* gene therapy deleivery of ChMIrp polypeptide is also envisioned. One example of a gene therapy technique is to use the ChMIrp gene (either genomic DNA, cDNA, and/or synthetic DNA encoding a ChMIrp polypeptide, or a fragment, variant, or derivative thereof) which may be operably linked to a constitutive or inducible promoter to form a "gene therapy DNA construct". The promoter may be homologous or heterologous to the endogenous ChMIrp gene, provided that it is active in the cell or tissue type into which the construct will be inserted. Other components of the gene therapy DNA construct may optionally include, DNA molecules designed for site-specific integration (*e.g.*, endogenous sequences useful for homologous recombination), tissue-specific promoter, enhancer(s) or silencer(s), DNA molecules capable of providing a selective advantage over the parent cell, DNA molecules useful as labels to identify transformed cells, negative selection systems, cell specific binding agents (as, for example, for cell targeting) cell-specific internalization factors, and transcription factors to enhance expression by a vector as well as factors to enable vector manufacture.

A gene therapy DNA construct can then be introduced into cells (either *ex vivo* or *in vivo*) using viral or non-viral vectors. One means for introducing the gene therapy DNA construct is via viral vectors. Suitable viral vectors typically used in gene therapy for delivery of gene therapy DNA constructs by means of viral vectors described herein. Certain retroviral vectors, will deliver the DNA construct to the chromosomal DNA of the cells, and the DNA construct can integrate into the chromosomal DNA. Other vectors will function as episomes, and the gene therapy DNA construct will remain in the cytoplasm.

In yet other embodiments, regulatory elements can be included for the controlled expression of the ChMIrp gene in the target cell. Such elements are turned on in response to an appropriate effector. In this way, a therapeutic polypeptide can be expressed when desired. One conventional control means involves the use of small molecule dimerizers or rapalogs (as described in WO9641865 (PCT/US96/099486); WO9731898 (PCT/US97/03137) and WO9731899 (PCT/US95/03157) used to dimerize chimeric proteins which contain a small molecule-binding domain and a domain capable of initiating biological process, such as a DNA-binding protein or transcriptional activation protein. The dimerization of the proteins can be used to initiate transcription of the transgene.

An alternative regulation technology uses a method of storing proteins expressed from the gene of interest inside the cell as an aggregate or cluster. The gene of interest is expressed as a fusion protein that includes a conditional aggregation domain which results in the retention of the aggregated protein in the endoplasmic reticulum. The stored proteins are stable and inactive inside the cell. The proteins can be released, however, by administering a drug (*e.g.*, small molecule ligand) that removes the conditional aggregation domain and thereby specifically breaks apart the aggregates or clusters so that the proteins may be secreted from the cell. *See*, *Science 287*:816-817, and 826-830 (2000).

Other suitable control means or gene switches include, but are not limited to, the following systems. Mifepristone (RU486) is used as a progesterone antagonist. The binding of a modified progesterone receptor ligand-binding domain to the progesterone antagonist activates transcription by forming a dimer of two

transcription factors which then pass into the nucleus to bind DNA. The ligan-binding domain is modified to eliminate the ability of the receptor to bind to the natural ligand. The modified steroid hormone receptor system is further described in U.S. 5,364,791; WO9640911; and WO9710337.

Yet another control system uses ecdysone (a fruit fly steroid hormone) which binds to and activates an ecdysone receptor (cytoplasmic receptor). The receptor then translocates to the nucleus to bind a specific DNA response element (promoter from ecdysone-responsive gene). The ecdysone receptor includes a transactivation domain/DNA-binding domain/ligand-binding domain to initiate transcription. The ecdysone system is further described in U.S. 5,514,578; WO9738117; WO9637609; and WO9303162.

Another control means uses a positive tetracycline-controllable transactivator. This system involves a mutated tet repressor protein DNA-binding domain (mutated tet R-4 amino acid changes which resulted in a reverse tetracycline-regulated transactivator protein, *i.e.*, it binds to a tet operator in the presence of tetracycline) linked to a polypeptide which activates transcription. Such systems are described in U.S. Patent Nos. 5,464,758; 5,650,298 and 5,654,168.

Additional expression control systems and nucleic acid constructs are described in U.S. Patent Nos. 5,741,679 and 5,834,186, to Innovir Laboratories Inc.

In vivo gene therapy may be accomplished by introducing the gene encoding a ChMIrp polypeptide into cells via local injection of a ChMIrp nucleic acid molecule or by other appropriate viral or non-viral delivery vectors. (Hefti, *Neurobiology, 25*:1418-1435, 1994). For example, a nucleic acid molecule encoding a ChMIrp polypeptide may be contained in an adeno-associated virus (AAV) vector for delivery to the targeted cells (*e.g.*, Johnson, International Publication No. WO95/34670; International Application No. PCT/US95/07178). The recombinant AAV genome typically contains AAV inverted terminal repeats flanking a DNA sequence encoding a ChMIrp polypeptide operably linked to functional promoter and polyadenylation sequences.

Alternative suitable viral vectors include, but are not limited to, retrovirus, adenovirus, herpes simplex virus, lentivirus, hepatitis virus, parvovirus,

papovavirus, poxvirus, alphavirus, coronavirus, rhabdovirus, paramyxovirus, and papilloma virus vectors. U.S. Patent No. 5,672,344 describes an *in vivo* viral-mediated gene transfer system involving a recombinant neurotrophic HSV-1 vector. U.S. Patent No. 5,399,346 provides examples of a process for providing a patient with a therapeutic protein by the delivery of human cells which have been treated *in vitro* to insert a DNA segment encoding a therapeutic protein. Additional methods and materials for the practice of gene therapy techniques are described in U.S. Patent No. 5,631,236 involving adenoviral vectors; U.S. Patent No. 5,672,510 involving retroviral vectors; and U.S. 5,635,399 involving retroviral vectors expressing cytokines.

Nonviral deleivery methods include, but are not limited to, liposome-mediated transfer, direct injection of naked DNA, receptor-mediated transfer (ligand-DNA complex), electroporation, calcium phosphate precipitation, and microparticle bombardment (*e.g.*, "gene gun"). Gene therapy materials and methods may also include the use of inducible promoters; tissue-specific enhancer-promoters, DNA sequences designed for site-specific integration, DNA sequences capable of providing a selective advantage over the parent cell, labels to identify transformed cells, negative selection systems and expression control systems (safety measures), cell-specific binding agents (for cell targeting), cell-specific internalization factors, and transcription factors to enhance expression by a vector as well as methods of vector manufacture. Such additional methods and materials for the practice of gene therapy techniques are described in U.S. Patent No. 4,970,154 involving electroporation techniques; WO96/40958 involving nuclear ligands; U.S. Patent No. 5,679,559 describing a lipoprotein-containing system for gene delivery; U.S. Patent No. 5,676,954 involving liposome carriers; U.S. Patent No. 5,593,875 concerning methods for calcium phosphate transfection; and U.S. Patent No. 4,945,050 wherein biologically active particles are propelled at cells at a speed whereby the particles penetrate the surface of the cells and become incorporated into the interior of the cells.

It is also contemplated that ChMIrp gene therapy or cell therapy can further include the delivery of one or more additional polypeptide(s) in the same or a

different cell(s). Such cells may be separately introduced into the patient, or the cells may be contained in a single implantable device, such as the encapsulating membrane described above, or the cells may be separately modified by means of viral vectors.

Another means to increase endogenous ChMIrp polypeptide expression in a cell via gene therapy is to insert one or more enhancer elements into the promoter of the ChMIrp gene, where the enhancer element(s) can serve to increase transcriptional activity of the ChMIrp polypeptides gene. The enhancer element(s) used will be selected based on the tissue in which one desires to activate the gene(s); enhancer elements known to confer promoter activation in that tissue will be selected. For example, if a gene encoding a ChMIrp polypeptide is to be "turned on" in T-cells, the *lck* promoter enhancer element may be used. Here, the functional portion of the transcriptional element to be added may be inserted into a fragment of DNA containing the ChMIrp polypeptide promoter (and optionally, inserted into a vector and/or 5′ and/or 3′ flanking sequence, etc.) using standard cloning techniques. This construct, known as a "homologous recombination construct", can then be introduced into the desired cells either *ex vivo* or *in vivo*.

Gene therapy also can be used to decrease ChMIrp polypeptide expression by modifying the nucleotide sequence of the endogenous promoter(s). Such modification is typically accomplished via homologous recombination methods. For example, a DNA molecule containing all or a portion of the promoter of the ChMIrp gene(s) selected for inactivation can be engineered to remove and/or replace pieces of the promoter that regulate transcription. For example, the TATA box and/or the binding site of a transcriptional activator of the promoter may be deleted using standard molecular biology techniques; such deletion can inhibit promoter activity thereby repressing the transcription of the corresponding ChMIrp gene. The deletion of the TATA box or transcription activator binding site in the promoter may be accomplished by generating a DNA construct comprising all or the relevant portion of the ChMIrp polypeptide promoter(s) (from the same or a related species as the ChMIrp gene(s) to be regulated) in which one or more of the TATA box and/or transcriptional activator binding site nucleotides are mutated via substitution, deletion and/or insertion of one or more nucleotide. As a result, the TATA box and/or

activator binding site has decreased activity or is rendered completely inactive. The construct, will typically contain at least about 500 bases of DNA that correspond to the native (endogenous) 5′ and 3′ DNA sequences adjacent to the promoter segment that has been modified. The construct may be introduced into the appropriate cells (either *ex vivo* or *in vivo*) either directly or via a viral vector as described herein. Typically, the integration of the construct into the genomic DNA of the cells will be via homologous recombination, where the 5′ and 3′ DNA sequences in the promoter construct can serve to help integrate the modified promoter region via hybridization to the endogenous chromosomal DNA.

Additional Uses of ChMIrp Nucleic Acids and Polypeptides

Nucleic acid molecules of the present invention (including those that do not themselves encode biologically active polypeptides) may be used to map the locations of the ChMIrp gene and related genes on chromosomes. Mapping may be done by techniques known in the art, such as PCR amplification and in situ hybridization.

ChMIrp nucleic acid molecules (including those that do not themselves encode biologically active polypeptides), may be useful as hybridization probes in diagnostic assays to test, either qualitatively or quantitatively, for the presence of a ChMIrp DNA or corresponding RNA in mammalian tissue or bodily fluid samples. ChMIrp may serve as a diagnoses/prognosis marker or assay for a wide variety of human cancers. Monitoring changes in the expression of ChMIrp during cancer treatment may be used as a surrogate marker to monitor tumor growth and treatment success.

The ChMIrp polypeptides may be used (simultaneously or sequentially) in combination with one or more cytokines, growth factors, antibiotics, anti-inflammatories, and/or chemotherapeutic agents as is appropriate for the indication being treated. ChMIrp may be useful as a small molecule inhibitor. In addition, peptide inhibitors designed from ChMIrp polypeptide may be used as a therapeutic or identifying substance which modulates ChMIrp polypeptide activity.

Other methods may also be employed where it is desirable to inhibit the activity of one or more ChMIrp polypeptides. Such inhibition may be effected by nucleic acid molecules which are complementary to and hybridize to expression control sequences (triple helix formation) or to ChMIrp mRNA. For example, antisense DNA or RNA molecules, which have a sequence that is complementary to at least a portion of the selected ChMIrp gene(s) can be introduced into the cell. Antisense probes may be designed by available techniques using the sequence of ChMIrp polypeptide disclosed herein. Typically, each such antisense molecule will be complementary to the start site (5' end) of each selected ChMIrp gene. When the antisense molecule then hybridizes to the corresponding ChMIrp mRNA, translation of this mRNA is prevented or reduced. Antisense inhibitors provide information relating to the decrease or absence of a ChMIrp polypeptide in a cell or organism.

Alternatively, gene therapy may be employed to create a dominant-negative inhibitor of one or more ChMIrp polypeptides. In this situation, the DNA encoding a mutant polypeptide of each selected ChMIrp polypeptide can be prepared and introduced into the cells of a patient using either viral or non-viral methods as described herein. Each such mutant is typically designed to compete with endogenous polypeptide in its biological role. Particularly, ChMIrp contains a kinase domain that may be useful in designing dominant negative gene therapy for treatment in a wide variety of tumors

In addition, a ChMIrp polypeptide, whether biologically active or not, may be used as an immunogen, that is, the polypeptide contains at least one epitope to which antibodies may be raised. Selective binding agents that bind to a ChMIrp polypeptide (as described herein) may be used for *in vivo* and *in vitro* diagnostic purposes, including, but not limited to, use in labeled form to detect the presence of ChMIrp polypeptide in a body fluid or cell sample. The antibodies may also be used to prevent, treat, or diagnose a number of diseases and disorders, including those recited herein. The antibodies may bind to a ChMIrp polypeptide so as to diminish or block at least one activity characteristic of a ChMIrp polypeptide, or may bind to a polypeptide to increase at least one activity characteristic of a ChMIrp polypeptide

(including by increasing the pharmacokinetics of the ChMIrp polypeptide).

The subject matter of the present invention is further described by following examples, which are intended for illustration purposes only, and should not be construed as limiting the scope of the invention in any way.

EXAMPLE 1

Isolation of a Murine cDNA Encoding a Chondromodulin-I Related Gene

A murine cDNA library was generated from total RNA extracted from the osteopetrotic bones of Osteoprotegerin transgenic mice (Simonet *et al.*, *Cell*, 89: 309-319, 1997) using a commercial RNA extraction kit (Pharmacia Biotech, Piscataway, NJ) according to the manufacturer's instuctions. Poly A^+ RNA was selected using Dynabeads Oligo (dT)25 columns (Dynal, Oslo, Norway). The cDNA was synthesized using the Superscript Plasmid System for cDNA synthesis and plasmid cloning (Gibco-BRL, Rockville, MD) according to the manufacturer's protocol. The resulting cDNA was digested with SalI and NotI restriction enzymes (Boehringer Mannheim, Indianapolis, IN) to create sticky ends to assist in ligation to a vector. The digested cDNA was ligated into a SALI/NOTI pre-digested pSPORT vector (Gibco-BRL) with T4 DNA ligase (Promega, Madison, WI). The ligated product was transformed into *E. coli* DB10B electrocompetent bacteria (Gibco-BRL) by electroporation. The transformed bacteria were plated onto LB agarose plates containing 100 mg/ml of ampicillin. The clones were selected randomly for sequencing.

Sequencing of clones, generated from the murine cDNA library described above, identified a DNA corresponding to an EST sequence, smbo2-00029-h3, which is now referred to as chondromodulin-I related peptide (ChMIrp). BLAST analysis of the SWISS-PROT database determined that the murine ChMIrp cDNA, set forth as SEQ ID NO: 3, encoded a protein exhibiting 32% identity over 116 amino acids to the amino-terminal portion of bovine chondromodulin-I polypeptide (SEQ ID NO: 7). The homology to chondromodulin-I allowed for the determination of both strands of the entire ChMIrp insert. As shown in Figure 1, the

polynucleotide sequence of murine ChMIrp (SEQ ID NO: 3) includes an open reading frame of 951 nucleotides. The murine ChMIrp polynucleotide sequence was deposited with the American Type Culture Collection (10801 University Boulevard Manassas, VA) on August 8, 2000 in compliance with the Budapest Treaty and given ATCC accession no. PTA-2329.

EXAMPLE 2

Identification of a Human Ortholog of Murine ChMIrp

BLAST analysis of the Genebank EST database with the full-length murine ChMIrp nucleotide sequence (SEQ ID NO: 3), revealed 7 human ESTs (Genebank accession numbers aa297231, t121280, t12179, ai123839, ai146280, 11453695, and ai147044) which encoded a human ortholog of ChMIrp. The full length human cDNA was generated by 3′ and 5′ RACE with the primers set out below in Table 1. The primer sequences were based on a consensus sequence derived from comparison of the 7 human ESTs.

Table III

Primer	Sequence	SEQ ID NO:
2244-23	CAC GAA GTA GAT GCC AGT GTA TCC	11
2244-24	GTG TAC TTC CAA TGT TTC ATC AGT GC	12
2244-19	CCA GTT ACA AGG CAT GAT GAC ACG	13
2244-20	CGT CCT CCT TGG TAG CAG TAT GG	14
AP-1	CCA TCC TAA TAC GAC TCA CTA TAG GGC	15
AP-2	ACT CAC TAT AGG GCT CGA GCG GC	16
2311-20	GTC AGT GAT TTG GGT CCC AGC AG	17
2311-21	CGT GAC CAT GTA TTG GAT CAA TCC C	18

Human Skeletal Muscle Marathon Ready cDNA and Human Heart Marathon ready cDNA were used as template DNA (Clontech, Palo Alto, CA) for 5′ RACE since Northern blot analysis (see Example 3 below) detected ChMIrp expression in human skeletal muscle and human heart. 5′ RACE from the Human Skeletal Muscle Marathon Ready cDNA was generated with the AP-1 Primer (SEQ ID NO: 15) as the 5′ PCR primer and Primer 2244-19 (SEQ ID NO: 13) as the 3′

PCR primer. 5′ RACE from Human Heart Marathon Ready cDNA was generated with AP1 as the 5′ PCR primer (SEQ ID NO: 15) and Primer 2244-23 as the 3′ PCR primer (SEQ ID NO: 11). The first round RACE reaction was carried out according to manufacturer's protocol (Clontech). One microliter of a 1:50 dilution of the first 5′ RACE product was used as a template for the nested 5′ RACE. The nested 5′ RACE reaction primers for skeletal muscle were AP-2 (SEQ ID NO: 16) as the 5′ primer and Primer 2244-20 (SEQ ID NO: 14) as the 3′ primer. In human heart nested 5′ RACE, AP-2 (SEQ ID NO: 16) was used as the 5′ PCR nested primer and 2244-22 was used as the 3′ PCR nested primer. The nested PCR reaction was carried out according to manufacturer's protocol (Clontech).

Human Skeletal Muscle Marathon Ready cDNA was used as template DNA for 3′ RACE. The 3′ RACE product was generated with Primer 2311-20 (SEQ ID NO: 14) as the 5′ primer and AP-1 (SEQ ID NO: 15) as the 3′ PCR primer. The first 3′ RACE reaction was carried out according to manufacturer's protocol (Clontech). A microliter of a 1:50 dilution of the first 3′ RACE product was used as template DNA for the nested 3′ RACE reaction. Primer 2311-21 (SEQ ID NO: 18) was used as the 5′ PCR nested primer, and the AP-2 (SEQ ID NO: 16) was used as the 3′ PCR nested primer. The nested PCR reaction was performed according to manufacturer's protocol (Clontech).

The products of the 5′ and 3′ RACE reactions were separated by electrophoresis on a 1% agarose gel. The appropriate sized bands were excised from the agarose gel and purified by Qiaquick gel extraction kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. The extracted DNA was incubated with Taq polymerase (Boehringer Mannheim) in 10 mM Tris-HCl (pH 8.3), 1.5 mM $MgCl_2$, 50 mM KCl, 0.2 mM dNTP's (dATP, dTTP, dGTP. dCTP) at 37°C for 5 min. The DNA was then ligated into pCR2.1 vector (Invitrogen, Carlsbad, CA) and transformed into bacteria according to the manufacturer's instructions. The size of the inserts were determined by EcoRI restriction enzyme (Boehringer Mannheim) digestion followed by electrophoresis on a 1% agarose gel. The identity of the 5′ and 3′ RACE products expressed by the pCR2.1 vector were verified by sequencing.

As shown in Figure 2, the human ChMIrp cDNA (SEQ ID NO: 1) consists of an open reading frame of 953 nucleotides (SEQ ID NO: 2), in addition to 86 bp in the 5′ untranslated region and 182 bp in the 3′ untranslated region. Alignment (Figure 3) of the human (huChIrp) and mouse (muChIrp) amino acid sequences (SEQ ID NOS: 2 and 4, respectively) exhibits 97% identity. The human ChMIrp polynucleotide sequence was deposited with the American Type Culture Collection (10801 University Boulevard Manassas, VA) on August 8, 2000 in compliance with the Budapest Treaty and given ATCC accession no. PTA-2328.

As shown in Figure 4, huChMIrp and muChMIrp were compared to the sequences of chondromodulin-I from several species including mouse (Genebank accession NO: U43509), rat (Genebank accession NO: AF051425) bovine (Swiss-Prot accession NO: CHM1_BOVIN), human (Genbank accession NO: AB006000), and rabbit (Genebank accession NO: AF072129) (SEQ ID NOS: 5-9, respectively). BLAST analysis determined that the muChMIrp open reading frame exhibited 39% identity over 289 amino acids of rat chondromodulin-I. In addition, the matches with human, mouse, bovine and rabbit chondromodulin-I exhibited about 35% identity. BLAST analysis determined that the open reading frame of muChMIrp exhibits 41% identity over 262 amino acids for rat, 40% identity over 262 amino acid for mouse, 36% identity over 314 amino acids for bovine, 37% identity over 262 amino acids for rabbit, and 37% identity over 260 amino acids for human chondromodulin-I polypeptides. In addition, murine ChMIrp exhibits 39% identity over 289 amino acids for chicken chondromodulin-I polypeptide (not shown). Since the conservation between human and mouse ChMIrp is so high, BLAST analysis data between huChMIrp and chondromodulin-I of other species would be comparable to muChMIrp identify described above.

EXAMPLE 3

Tissue Specific Expression of ChMIrp

Tissue-specific expression patterns of the ChMIrp gene were investigated by Northern Blot. A PCR-generated ^{32}P-labeled probe was used to detect the presence of ChMIrp transcripts in various tissues from both human and mouse.

The mouse ChMIrp cDNA (SEQ ID NO: 3) inserted into pSPORT vector (Gibco-BRL) was used as a template to PCR amplify the entire coding region to use as a probe for Northern blot analysis. The 5′ primer was defined as Primer 2245-78 and consisted of the sequence ATGGCAAAGAATCCTCCAGAGAAC (SEQ ID NO: 19) and the 3′ primer was designated Primer 2245-79 and consisted of the sequence CTATTAGACTCTCCCAAGCATGCG (SEQ ID NO: 20). The PCR reaction was performed in a 100 ml volume containing 10 mM Tris-HCl (pH 8.3), 1.5 mM $MgCl_2$, 50 mM KCl, 0.2 mM of dATP, dTTP and dGTP, 0.01 mM dCTP, 0.17 mM (α-^{32}P)dCTP, 0.4 mM of each primer and 10 ng of muChMIrp template DNA. The PCR parameters consisted of a denaturing step of 94°C for 2 minutes cycling 40 times at 94°C for 30 seconds, annealing at 70°C for 30 seconds, and extension at 72°C for 1 minute. The probe was purified over a Sepharose G50 column (5′-3′ Inc, Boulder, CO).

Commercially available multiple tissue Northern blots (Clontech, Palo Alto, Ca) containing human, embryonic mouse and adult mouse tissues and a commercially available Zoo Blot (Clontech) were probed with muChMIrp. In addition, RNA was isolated from the femurs and tibias of Osteoprotegerin (OPG) transgenic mice, OPG knockout mouse (Bucay *et al.*, *Genes Dev.*, 12: 1260-1268) and normal CD-1 mice by standard techniques (Sambrook *et al.*, *Molecular Cloning*, Cold Springs Harbor Laboratory Press, New York, 1989). Tissues were lysed with 20 ml of TRIzol reagent (Gibco-BRL), homogenized for 30 seconds, and extracted with 4 ml of chloroform. The tubes were centrifuged at 4000 rpm for 30 minutes, and the aqueous phase was transferred to a new tube. RNA was precipitated by adding 10 ml isopropanol, mixing, and centrifuging for 30 minutes at 4200 rpm. The RNA pellet was washed with 10 ml of 70% ethanol, dried briefly and resuspended in 0.5 ml TE buffer. RNA was fractionated using a formaldehyde/agarose gel electrophoresis system. Following electrophoresis, the gel was processed and the RNA transferred to a nylon membrane (*See* Sambrook *et al., supra*).

Northern blots were prehybridized in Rapid-hyb buffer (Amersham Life Sciences, Arlington Heights, IL) at 65°C for 30 minutes. The blots were then hybridized at 65°C for 2 hours in the same solution with the addition of the ^{32}P-

antisense RNA probe complementary to full length muChMIrp and with sense (control) probes overnight at 55°C. The antisense and sense ^{33}P-labeled probes were obtained by *in vitro* transcription of plasmid DNA containing muChMIrp cDNA. Following hybridization, sections were washed 2 times in 4x SSC at 55°C, treated with 20 mg/ml RNase A to remove unhybridized probe, then subjected to a high stringency wash in 0.1x SSC at 55°C. Slides were dipped in Kodak NTB2 emulsion, exposed at 55°C for two to three weeks, developed, and then counterstained. Sections were examined with darkfield and standard illumination to allow simultaneous evaluation of tissue morphology and hybridization signal. The following tissues were then examined: brain, parotid, submandibular and sublingual glands, esophagus, stomach, duodenum, jejunum, ileum, proximal and distal colon, liver, pancreas, heart, lung, trachea, blood vessels, lymph nodes, spleen, thymus, bone marrow, kidney, bladder, thyroid gland, adrenal gland, testis, prostate, ovary, uterus, oviduct, placenta, bone, skeletal muscle, skin, and adipose tissue.

The muChMIrp antisense probe produced a clear signal detectable above a very low level of background seen with the sense control probe. No signal was detected in E8.5, 10.5 or 11.5 sections. From E13.5 through adult sections expression was detected in tendon and possibly fascia. While there was clear signal present in all identifiable tendons, no signal was detected in muscle, bone or cartilage. Signal was also detected in cells adjacent to hair follicles in the skin and in specialized epithelial cells overlying lymphoid patches in the intestine known as M-cells. Additional sites of expression were detected in the thymic medulla, cerebral cortex just above the corpus callosum in the brain, and granulosa cells surrounding the developing follicles in the ovary.

EXAMPLE 4

ChMIrp Genomic DNA Characterization

A BLAST analysis identified Genbank assession number AL035608 to contain the human ChMIrp genomic sequence. The sequence was from clone 479J7 that contains human genomic DNA from chromosome Xq21.33-23. The human

ChMIrp genomic DNA (SEQ ID NO: 10) contains seven exons and six introns and is presented in Figure 5.

EXAMPLE 5

Production of ChMIrp polypeptides

A. Expression of ChMIrp Polypeptide in Bacteria

PCR is used to amplify template DNA sequences encoding the ChMIrp polypeptide using primers corresponding to the 5′ and 3′ ends of the sequence. The amplified DNA products may be modified to contain restriction enzyme sites to allow insertion into an expression vector using standard recombinant DNA methodology. An exemplary vector, such as pAMG21 (ATCC NO: 98113) containing the lux promoter and a gene encoding kanamycin resistance is digested with BamHI and NdeI for directional cloning of inserted DNA. The ligated mixture is transformed into *E. coli* host strain Top 10 (Invitrogen) by electroporation and transformants selected for kanamycin resistance. Plasmid DNA from selected colonies is isolated and subjected to sequencing to confirm the presence of the insert.

Transformed host cells are incubated in 2x YT medium containing 30 mg/ml kanamycin at 30°C prior to induction. Gene expression is induced by addition of N-(3-oxohexanoyl)-Dl-homoserine lactone to a final concentration of 30 ng/ml followed by incubation at either 30°C or 37°C for six hours. Expression of ChMIrp polypeptide is evaluated by centrifugation of the culture, resuspension and lysis of the bacterial pellets, and analysis of host cell proteins by SDS polyacrylamide gel electrophosis (SDS-PAGE).

Inclusion bodies containing ChMIrp polypeptide are purified as follows: Bacterial cells are pelleted by centrifugation and resuspended in water. The cell suspension is lysed by sonication and pelleted by centrifuation at 195,000 x g for 5 to 10 minutes. The supernate is discarded and the pellet washed and transferred to a homogenizer. The pellet is homogenized in 5 ml of a Percoll solution (75% liquid Percoll/0.15 M NaCl) until uniformly suspended and then diluted and centrifuged at 21,600 x g for 30 minutes. Gradient fractions containing the inclusion bodies are

recovered and pooled. The isolated inclusion bodies are then analyzed by SDS-PAGE.

A single band on a SDS-PAGE gel corresponding to *E. coli* produced ChMIrp polypeptide is excised from the gel and the N-terminal amino acid sequence is determined essentially as described by Matsudaira *et al.* (*J. Biol. Chem.*, 262: 10-35, 1987).

B. Expression of ChMIrp Polypeptide in Mammalian Cells

PCR is used to amplify template DNA sequences encoding the ChMIrp polypeptide using primers corresponding to the 5′ and 3′ ends of the sequence. The amplified DNA products may be modified to contain restriction enzyme sites to allow for insertion into expression vectors. PCR products are gel purified and inserted into expression vectors using standard recombinant DNA methodology. An exemplary vector, pCEP4 (Invitrogen), which contains an Epstein-Barr virus origin of replication, may be used for expression of ChMIrp in COS cells. Amplified and gel purified PCR products are ligated into pCEP4 and lipofected into COS cells. The transfected cells are selected in 100 mg/ml hygromycin and the resulting drug-resistant cultures are grown to confluence. The cells are then cultured in serum-free media for 72 hours, the conditioned media removed and ChMIrp polypeptide expression analyzed by SDS-PAGE.

ChMIrp polypeptide expression may be detected, for example, by silver staining. Alternatively, ChMIrp is produced as a fusion protein with an epitope tag, such as an IgG constant domain or a FLAG epitope, which may be detected by Western blot analysis using antibodies to the tag peptide.

ChMIrp polypeptides may be excised from a SDS-PAGE gel, or ChMIrp fusion proteins are purified by affinity chromatography to the epitope tag, and subjected to N-terminal amino acid sequence analysis, as described above.

EXAMPLE 6

Production of ChMIrp Antibodies

Antibodies to ChMIrp polypeptides may be obtained by immunization with purified protein or with ChMIrp peptides produced by biological or chemical

synthesis. Substantially pure ChMIrp polypeptide or polypeptides may be isolated from transfected cells as described in Example 5. Concentration of protein in the final preparation may be adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibodies to the proteins can then be prepared by any of the procedures known in the art for generating antibodies, such as those described in Hudson and Bay, (*Practical Immunology*, 2nd Ed., Blackwell Scientific Productions).

A. Anti-ChMIrp Monoclonal Antibody Production

A monclonal antibody to an epitope of any of the peptides identified and isolated as described can be prepared from a murine hybridoma according to the classical methods of Kohler and Milstein (*Nature*, 256: 495, 1975) or derivative methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells such as NS-1 cells, and the excess unfused cells destroyed by growth of the system on selective media comprising hypoxanthine; aminopterin; and thymidine (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. After selection, tissue culture supernatants are taken from each fusion well and tested for ChMIrp antibody production by EIA. Selective positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis *et al.* (*Basic Methods in Molecular Biology*, Section 21-2, Elsevier, New York, NY).

B. Polyclonal ChMIrp Antibody Production

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein as described above. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small

molecules tend to be less immunogenic than large molecules and may require the use of carrier adjuvants. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng levels) of antigen administered at multiple intradermal sites appear to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis *et al.* (*J. Clin. Enodcrinol. Metab.*, 33: 988-991, 1971).

Booster injections can be given at regular intervals, and anti-serum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. *See* for example, Ouchterlony *et al.* (Chap.19 in: *Handbook of Experimental Immunology*, D. Weir (ed.), Blackwell, 1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 mM). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fischer, D., (Chap. 42 In: *Manual of Clinical Immunology*, 2nd Ed. (Rose and Friedman, eds.) Amer. Soc. For Microbiol., Washington, D.C., 1980).

Alternative procedures for obtaining anti-ChMIrp antibodies may also be employed, such as immunization of transgenic mice harboring human IgG loci for production of fully human antibodies, and screening of synthetic antibody libraries, such as those generated by mutagenesis of an antibody variable domain.

EXAMPLE 7

Functional Analysis of the Role of ChMIrp

To determine the functional role of ChMIrp *in vivo*, the ChMIrp gene is either over-expressed in the germ line of animals or inactivated in the germ line of mammals by homologous recombination. Animals in which the gene is over-expressed under the regulatory control of exogenous or endogenous promoter elements are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are also known as "knockout" animals. Exemplary mammals include rabbits and rodent species such as mice.

Exemplary procedures are described in U.S. Patent No. 5,489,743 and International Patent Publication No. WO 94/28122.

Transgenic animals allow for the determination of the effect(s) of over expression or inappropriate expression of the ChMIrp on development and disease processes. ChMIrp transgenic animals can also serve as a model system to test compounds that can modulate receptor activity.

The "knockout" animals allow for the determination of the role of ChMIrp polypeptide in embryonic development, and in immune and proliferative responses. The role of ChMIrp polypeptide in development, and proliferative responses is determined by analyzing the effects of gene knockout on the development of the embryo as well as on the development and differentiation of various organs and tissues for example skeletal components such as bones and cartilage in these animals.

EXAMPLE 8

Biological Activity of ChMIrp Polypeptide

Proliferation and morphological assays may be carried out to determine if ChMIrp biological activity is similar to that of chondromodulin-I. For all assays, recombinant ChMIrp, from COS cells prepared and purified as described in Example 5, is employed.

A. Stimulation of Chrondocytes

The ability of ChMIrp to stimulate DNA and proteoglycan synthesis in chondrocytes may be investigated as a chondromodulin family activity. Chondrocytes are isolated from the growth plate cartilage of the rib from young rabbits as described by Shimomura *et al.* (*Calcif. Tissue Res.* 19: 179-187, 1975). The isolated chondrocytes are plated at 1 x 10^4 cells/well in 96-well microtiter plates and cultured in modified Eagle's medium (MEM) supplemented with 10% fetal bovine serum (FBS). Upon confluence, the proteoglycan synthesis is measured as described in Hiraki *et al.* (*Eur. J. Biochem.*, 260: 869-878, 1999). Briefly, the chondrocytes are pre-incubated in 0.3% serum. After 24 hours, the medium is replaced with that containing 0.3% serum and recombinant ChMIrp polypeptide (10-

1000 ng/ml). After 3 hours, the cells are labeled with 5 mCi/ml [^{35}S]-sulphate for an additional 17 hours. Subsequently, the proteoglycans are precipitated by 1% cetylpyridinium chloride, and the incorporated radioactivity is measured. Insulin-like growth factor-I or -II may be used as positive controls.

To determine if DNA synthesis is stimulated by ChMIrp polypeptide, the isolated chondrocytes are plated at 1 x 10^4 cells/well in a 96-well microtiter plate in MEM medium containing 10% FBS. After 24 hours, the medium is changed to that containing 0.3% serum plus recombinant ChMIrp polypeptide (10-1000 ng/ml) and the cells are incubated for 20 hours. The cells are then labeled with [^{3}H]-thymidine (5 mCi/ml) for an additional 4 hours. Subsequently, the cells are washed with PBS and fixed with cold 100% methanol for 10 minutes. After fixation, the incorporated radioactivity is precipitated with 10% trichloroacetic acid and counted.

Chondromodulin-I synergizes with FGF-2 to induce chondrocyte colony formation in soft agar. To determine if ChMIrp polypeptide exhibits this activity, isolated chondrocytes (5 x 10^3 cells/well) are suspended in 0.5 ml of 0.41% agarose in HAM's F-12 medium supplemented with 5% FBS, 0.2 mM hydrocortisone, and 60 mg/ml transferrin as described in Inoue *et al.* (*Biochem. Biophys. Res. Comm.*, 241: 395-400, 1997). This cellular suspension is poured over a base layer of 0.72% agarose and incubated overnight. Increasing concentrations of recombinant ChMIrp polypeptide (1-1000 ng/ml) and 1 ng/ml FGF-2 are diluted in serum-free F-12 medium containing 0.5% bovine serum albumin. This mixture is added evenly to the top layer of the wells. After 10 days, colonies are counted under a phase-contrast microscope.

B. Endothelial Cell Stimulation

Endothelial cell growth and morphogenesis in the presence of ChMIrp polypeptide is assayed to determine if inhibition of endothelial cell growth and tube formation occurs similar to that in the presence of chondromodulin-I. The growth and tube morphogenesis in bovine carotid artery endothelial cells *in vitro* are measured as described by Hiraki *et al.* (*FEBS*, 415: 321-324, 1997). Briefly, bovine carotid artery (BCAE) cells are isolated by gently scraping the intimal surface of the

carotid artery. The isolated BCAE cells are cultured and expanded in RPMI-1640 supplemented with 10% FBS. To determine if recombinant ChMIrp can inhibit proliferation of BCAE cells, [^{3}H]-thymidine incorporation is measured as described above. To examine ChMIrp effect on endothelial cell tube morphogenesis, the BCAE cells (1×10^5) cells/well are grown in 12-well plates containing in 0.3% type I collagen gel diluted in 0.1 M NaOH and 10x MEM media. After 24 hours, the medium is aspirated and an aqueous mixture (70 ml) containing recombinant ChMIrp polypeptide (10-1000 ng/ml) is added. A collagen solution is then overlaid to create the top layer. After 3 days, cellular morphological changes into tube-like networks are observed under a phase-contrast microscope. An *in vivo* assay to measure endothelial cell vascularization is the chicken chorioallantoic membrane assay as described by Hiraki *et al.* (*Eur. J. Biochem.* 260: 869-878, 1999). Briefly, fertilized white Leghorn chicken eggs are incubated at 37.8°C. On day 5, the air chamber is punctured and a 1 cm^2 window is cut into the egg. The recombinant ChMIrp polypeptide (500 ng/ml) is diluted into 0.75% agarose. The solidified gel is placed on the choriallantoic membrane within the egg for 24 hours. The membranes are investigated under a dissecting microscope for fine capillary formation.

C. Osteoblast Stimulation

The effect of ChMIrp on osteoblast proliferation can also be measured as an indicator of chondromodulin-I activity as described by Mori *et al.* (*FEBS* 406: 310-314, 1997). The clonal osteoblast cell line, MC3T3-E1, is treated with increasing concentrations of recombinant ChMIrp (10-1000 ng/ml). As a measure of DNA synthesis, the osteoblast [^{3}H]-thymidine incorporation is measured as described above.

The envisioned biological function for ChMIrp polypeptide is similar to those of the chondromodulin-I. Chondromodulin-I, among other things, is known to stimulate the growth and differentiation of chondrocytes. Furthermore, chondromodulin-I has anti-angiogenic activity since it inhibits endothelial cell proliferation and fine capillary formation, *in vivo*. As such, ChMIrp may play a role in cartilage development and blood vessel formation.

It has been determined by Northern blot analysis and *in situ* hybridization that the ChMIrp polypeptide is expressed in the tendons, skeletal muscle, thymus, ovary, cerebral cortex in the brain, M cells in the intestine, and cells adjacent to the hair follicle. Expression in tendon and muscle indicate that ChMIrp may play a role in the development of tendons and muscle and in attachment of muscle to bone. Expression in the thymus, hair follicle and M cells of the intestine indicate a possible role of ChMIrp in immune function. ChMIrp may act as a growth factor involved in the regeneration (growth and development) of tissues and specialized cell types present in the tendons, skeletal muscle, thymus, ovary, brain, intestine and hair follicle.

Based on these potential functions, ChMIrp may be useful for the diagnosis and/or treatment of tendon diseases (such as tendinitis and tendon tear), skeletal muscle diseases (such as cachexia and muscular dystrophy), immune system dysfunction diseases (such as inflammation and allergy, poor wound healing, arthritis and allergies), and infertility diseases.

While the present invention has been described in terms of the preferred embodiments, it is understood that variations and modifications will occur to those skilled in the art. Therefore, it is intended that the appended claims cover all such equivalent variations which come within the scope of the invention as claimed.

WHAT IS CLAIMED

1. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

(a) the nucleotide sequence set forth in SEQ ID NO: 1;

(b) a nucleotide sequence encoding the polypeptide set forth in SEQ ID NO: 2;

(c) a nucleotide sequence which hybridizes under moderately or highly stringent conditions to the complement of (a) or (b), wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2; and

(d) a nucleotide sequence complementary to any of (a)-(c).

2. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence encoding a polypeptide that is at least about 70 percent identical to the polypeptide set forth in SEQ ID NO: 2, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(b) a nucleotide sequence encoding an allelic variant or splice variant of the nucleotide sequence set forth in SEQ ID NO: 1, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(c) a nucleotide sequence of SEQ ID NO: 1; (a); or (b) encoding a polypeptide fragment of at least about 25 amino acid residues, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(d) a nucleotide sequence of SEQ ID NO: 1, or (a)-(c) comprising a fragment of at least about 16 nucleotides;

(e) a nucleotide sequence which hybridizes under moderately or highly stringent conditions to the complement of any of (a)-(d), wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2; and

(f) a nucleotide sequence complementary to any of (a)-(c).

3. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence encoding a polypeptide set forth in SEQ ID NO: 2 with at least one conservative amino acid substitution, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(b) a nucleotide sequence encoding a polypeptide set forth in SEQ ID NO: 2 with at least one amino acid insertion, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(c) a nucleotide sequence encoding a polypeptide set forth in SEQ ID NO: 2 with at least one amino acid deletion, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(d) a nucleotide sequence encoding a polypeptide set forth in SEQ ID NO: 2 which has a C- and/or N- terminal truncation, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(e) a nucleotide sequence encoding a polypeptide set forth in SEQ ID NO: 2 with at least one modification selected from the group consisting of amino acid substitutions, amino acid insertions, amino acid deletions, C-terminal truncation, and N-terminal truncation, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(f) a nucleotide sequence of (a)-(e) comprising a fragment of at least about 16 nucleotides;

(g) a nucleotide sequence which hybridizes under moderately or highly stringent conditions to the complement of any of (a)-(f), wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2; and

(h) a nucleotide sequence complementary to any of (a)-(e).

4. A vector comprising the nucleic acid molecule of claims 1, 2, or 3.

5. A host cell comprising the vector of claim 4.

6. The host cell of claim 5 that is a eukaryotic cell.

7. The host cell of claim 5 that is a prokaryotic cell.

8. A process of producing a ChMIrp polypeptide comprising culturing the host cell of claim 5 under suitable conditions to express the polypeptide, and optionally isolating the polypeptide from the culture.

9. A polypeptide produced by the process of claim 8.

10. The process of claim 8, wherein the nucleic acid molecule comprises promoter DNA other than the promoter DNA for the native ChMIrp polypeptide operatively linked to the DNA encoding the ChMIrp polypeptide.

11. The isolated nucleic acid molecule according to claim 2 wherein the percent identity is determined using a computer program selected from the group consisting of GAP, BLASTP, BLASTN, FASTA, BLASTA, BLASTX, BestFit, and the Smith-Waterman algorithm.

12. A process for identifying candidate inhibitors of ChMIrp polypeptide activity or production comprising exposing a cell according to claims 5, 6, or 7 to the candidate inhibitors, and measuring ChMIrp polypeptide activity or production in said cell, comparing activity of ChMIrp in cells exposed to the candidate inhibitor with activity in cells not exposed to the candidate inhibitor.

13. A process for identifying candidate stimulators of ChMIrp polypeptide activity or production comprising exposing a cell according to claims 5, 6, or 7 to the candidate stimulators, and measuring ChMIrp polypeptide activity or production in said cell, comparing activity of ChMIrp in cells exposed to the candidate stimulator with activity in cells not exposed to the candidate stimulator.

14. An isolated polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 2.

15. An isolated polypeptide comprising the amino acid sequence selected from the group consisting of:

(a) the mature amino acid sequence set forth in SEQ ID NO: 2, comprising a mature amino terminus at residue 1, optionally further comprising an amino-terminal methionine;

(b) an amino acid sequence for an ortholog of SEQ ID NO: 2, wherein the encoded polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(c) an amino acid sequence that is at least about 70 percent identical to the amino acid sequence of SEQ ID NO: 2, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(d) a fragment of the amino acid sequence set forth in SEQ ID NO: 2 comprising at least about 25 amino acid residues, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(e) an amino acid sequence for an allelic variant or splice variant of either the amino acid sequence set forth in SEQ ID NO: 2, or at least one of (a)-(c) wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2.

16. An isolated polypeptide comprising the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth in SEQ ID NO: 2 with at least one conservative amino acid substitution, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(b) the amino acid sequence set forth in SEQ ID NO: 2 with at least one amino acid insertion, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(c) the amino acid sequence set forth in SEQ ID NO: 2 with at least one amino acid deletion, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2;

(d) the amino acid sequence set forth in SEQ ID NO: 2 which has a C- and/or N-terminal truncation, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2; and

(e) the amino acid sequence set forth in SEQ ID NO: 2, with at least one modification selected from the group consisting of amino acid substitutions, amino acid insertions, amino acid deletions, C-terminal truncation, and N-terminal truncation, wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2.

17. A polypeptide according to claim 15 or 16 wherein the amino acid at position 276 of SEQ ID NO: 2 is cysteine, serine or alanine.

18. A polypeptide according to claim 15 or 16 wherein the amino acid at position 280 of SEQ ID NO: 2 is cysteine, serine or alanine.

19. A polypeptide according to claim 15 or 16 wherein the amino acid at position 281 of SEQ ID NO: 2 glutamic acid or aspartic acid.

20. A polypeptide according to claim 15 or 16 wherein the amino acid at position 285 of SEQ ID NO: 2 is glycine, proline or alanine.

21. A polypeptide according to claim 15 or 16 wherein the amino acid at position 297 of SEQ ID NO: 2 is arginine, lysine, glutamine, or asparagine.

22. A polypeptide according to claim 15 or 16 wherein the amino acid at position 300 of SEQ ID NO: 2 is cysteine, serine or alanine.

23. A polypeptide according to claim 15 or 16 wherein the amino acid at position 306 of SEQ ID NO: 2 is cysteine, serine or alanine.

24. A polypeptide according to claim 15 or 16 wherein the amino acid at position 310 of SEQ ID NO: 2 is valine, isoleucine, methionine, leucine, phenylalanine, alanine or norleucine.

25. An isolated polypeptide encoded by the nucleic acid molecule of claims 1, 2, or 3.

26. The isolated polypeptide according to claim 15 wherein the percent identity is determined using a computer program selected from the group consisting of GAP, BLASTP, BLASTN, FASTA, BLASTA, BLASTX, BestFit, and the Smith-Waterman algorithm.

27. An antibody produced by immunizing an animal with a peptide comprising an amino acid sequence of SEQ ID NO: 2.

28. An antibody or fragment thereof that specifically binds the polypeptide of claims 14, 15, or 16.

29. The antibody of claim 28 that is a monoclonal antibody.

30. A hybridoma that produces a monoclonal antibody that binds to a peptide comprising an amino acid sequence of SEQ ID NO: 2.

31. A method of detecting or quantitating the amount of ChMIrp in a sample comprising contacting a sample suspected of containing ChMIrp polypeptide with the anti-h2520-109 antibody or fragment of claims 27, 28, 29 and detecting the binding of said antibody or fragment.

32. A selective binding agent or fragment thereof that specifically binds at least one polypeptide wherein said polypeptide comprises the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth in SEQ ID NO: 2;

(b) a fragment of the amino acid sequence set forth in at least one of SEQ ID NO: 2; and

(c) a naturally occurring variant of (a) or (b).

33. The selective binding agent of claim 32 that is an antibody or fragment thereof.

34. The selective binding agent of claim 32 that is a humanized antibody.

35. The selective binding agent of claim 32 that is a human antibody or fragment thereof.

36. The selective binding agent of claim 32 that is a polyclonal antibody or fragment thereof.

37. The selective binding agent claim 32 that is a monoclonal antibody or fragment thereof.

38. The selective binding agent of claim 32 that is a chimeric antibody or fragment thereof.

39. The selective binding agent of claim 32 that is a CDR-grafted antibody or fragment thereof.

40. The selective binding agent of claim 32 that is an antiidiotypic antibody or fragment thereof.

41. The selective binding agent of claim 32 which is a variable region fragment.

42. The variable region fragment of claim 41 which is a Fab or a Fab' fragment.

43. A selective binding agent or fragment thereof comprising at least one complementarity determining region with specificity for a polypeptide having the amino acid sequence of SEQ ID NO: 2.

44. The selective binding agent of claim 32 which is bound to a detectable label.

45. The selective binding agent of claim 32 which antagonizes ChMIrp polypeptide biological activity.

46. A method for treating, preventing, or ameliorating a disease, condition, or disorder comprising administering to a patient an effective amount of a selective binding agent according to claim 32.

47. A selective binding agent produced by immunizing an animal with a polypeptide comprising an amino acid sequence of SEQ ID NO: 2.

48. A hybridoma that produces a selective binding agent capable of binding a polypeptide according to claims 14, 15, or 16.

49. A composition comprising the polypeptide of claims 14, 15, or 16 and a pharmaceutically acceptable formulation agent.

50. The composition of claim 49 wherein the pharmaceutically acceptable formulation agent is a carrier, adjuvant, solubilizer, stabilizer, or anti-oxidant.

51. The composition of claim 50 wherein the polypeptide comprises the mature amino acid sequence set forth in SEQ ID NO: 2.

52. A polypeptide comprising a derivative of the polypeptide of claims 14, 15, or 16.

53. The polypeptide of claim 52 which is covalently modified with a water-soluble polymer.

54. The polypeptide of claim 53 wherein the water-soluble polymer is selected from the group consisting of polyethylene glycol, monomethoxy-polyethylene glycol, dextran, cellulose, poly-(N-vinyl pyrrolidone) polyethylene glycol, propylene glycol homopolymers, polypropylene oxide/ethylene oxide co-polymers, polyoxyethylated polyols, and polyvinyl alcohol.

55. A composition comprising a nucleic acid molecule of claims 1, 2, or 3 and a pharmaceutically acceptable formulation agent.

56. A composition of claim 55 wherein said nucleic acid molecule is contained in a viral vector.

57. A viral vector comprising a nucleic acid molecule of claims 1, 2, or 3.

58. A fusion polypeptide comprising the polypeptide of claims 14, 15, or 16 fused to a heterologous amino acid sequence.

59. The fusion polypeptide of claim 58 wherein the heterologous amino acid sequence is an IgG constant domain or fragment thereof.

60. A method for treating, preventing or ameliorating a medical condition in a mammal resulting from decreased levels of ChMIrp polypeptide comprising administering to a patient the polypeptide of claims 14, 15, or 16 or the polypeptide encoded by the nucleic acid of claims 1, 2, or 3 to said mammal.

61. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject caused by or resulting from abnormal levels of ChMIrp polypeptide comprising:

(a) determining the presence or amount of expression of the polypeptide of claims 14, 15, or 16 or the polypeptide encoded by the nucleic acid molecule of claims 1, 2, or 3 in a sample; and

(b) comparing the level of ChMIrp polypeptide in a biological, tissue or cellular sample from normal subjects or the subject at an earlier time, wherein susceptibility to a pathological condition is based on the presence or amount of expression of the polypeptide.

62. A device, comprising:

(a) a membrane suitable for implantation; and

(b) cells encapsulated within said membrane, wherein said cells secrete a polypeptide of claims 14, 15, or 16, and wherein said membrane is permeable to said protein and impermeable to materials detrimental to said cells.

63. A device, comprising:

(a) a membrane suitable for implantation; and

(b) the ChMIrp polypeptide encapsulated within said membrane, wherein said membrane is permeable to the polypepetide.

64. A method of identifying a compound which binds to a polypeptide comprising:

(a) contacting the polypeptide of claims 14, 15, or 16 with a compound; and

(b) determining the extent of binding of the polypeptide to the compound.

65. A method of modulating levels of a polypeptide in an animal comprising administering to the animal the nucleic acid molecule of claims 1, 2, or 3.

66. A transgenic non-human mammal comprising the nucleic acid molecule of claims 1, 2, or 3.

67. A diagnostic reagent comprising a detectably labeled polynucleotide encoding the amino acid sequence set out in SEQ ID NO: 2, or a fragment, variant or homolog thereof including allelic variants and spliced variants thereof.

68. The diagnostic reagent of claim 67, wherein said labeled polynucleotide is a first-strand cDNA.

69. A method for determine the presence of ChMIrp nucleic acids in a biological sample comprising the steps of:

(a) providing a biological sample suspected of containing ChMIrp nucleic acids;

(b) contacting the biological sample with a diagnostic reagent according to claim 60 under conditions wherein the diagnostic reagent will hybridize with h2520-109nucleic acids contained in said biological sample;

(c) detecting hybridization between ChMIrp nucleic acid in the biological sample and the diagnostic reagent; and

(d) comparing the level of hybridization between the biological sample and diagnostic reagent with the level of hybridization between a known concentration of ChMIrp nucleic acid and the diagnostic reagent.

70. A method for detecting the presence of ChMIrp nucleic acids in a tissue or cellular sample comprising the steps of:

(a) providing a tissue or cellular sample suspected of containing ChMIrp nucleic acids;

(b) contacting the tissue or cellular sample with a diagnostic reagent according to claim 68 under conditions wherein the diagnostic reagent will hybridize with ChMIrp nucleic acids;

(c) detecting hybridization between ChMIrp nucleic acid in the tissue or cellular sample and the diagnostic reagent; and

(d) comparing the level of hybridization between the tissue or cellular sample and diagnostic reagent with the level of hybridization between a known concentration of ChMIrp nucleic acid and the diagnostic reagent.

71. The method of claim 70 or 71 wherein said polynucleotide molecule is DNA.

72. The method of claim 70 or 71 wherein said polynucleotide molecule is RNA.

73. An antagonist of ChMIrp polypeptide activity selected from the group consisting of ChMIrp selective binding agents, small molecules, antisense oligonucleotides, and peptides or derivatives thereof having specificity for ChMIrp polypeptide.

Figure 1 Murine ChMIrp

```
                    10                  30                  50
SEQ ID NO:3 ATGGCAAAGAATCCTCCAGAGAACTGTGAGGGCTGTCACATTCTAAATGC
SEQ ID NO:4  M  A  K  N  P  P  E  N  C  E  G  C  H  I  L  N  A
                           70                  90
            AGAAGCTCTGAAATCTAAGAAGATATGTAAATCACTGAAGATTTGTGGAC
              E  A  L  K  S  K  K  I  C  K  S  L  K  I  C  G
                  110                 130                 150
            TAGTGTTTGGTATCCTGGCCTTAACTCTAATTGTCCTGTTTTGGGGGAGC
            L  V  F  G  I  L  A  L  T  L  I  V  L  F  W  G  S
                           170                 190
            AAACACTTCTGGCCCGAGGTATCCAAGAAAACCTATGACATGGAGCACAC
             K  H  F  W  P  E  V  S  K  K  T  Y  D  M  E  H  T
                  210                 230                 250
            TTTCTACAGCAACGGCGAGAAGAAGAAGATTTACATGGAAATTGATCCCA
              F  Y  S  N  G  E  K  K  K  I  Y  M  E  I  D  P
                           270                 290
            TAACCAGAACAGAAATATTCAGAAGTGGAAATGGCACTGATGAAACATTG
            I  T  R  T  E  I  F  R  S  G  N  G  T  D  E  T  L
                  310                 330                 350
            GAAGTCCATGACTTTAAAAATGGATACACTGGCATCTACTTTGTAGGTCT
             E  V  H  D  F  K  N  G  Y  T  G  I  Y  F  V  G  L
                           370                 390
            TCAAAAATGCTTTATTAAAACTCAAATCAAAGTGATTCCTGAATTTTCTG
              Q  K  C  F  I  K  T  Q  I  K  V  I  P  E  F  S
                  410                 430                 450
            AACCAGAGGAAGAAATAGATGAGAATGAAGAAATTACTACAACTTTCTTT
            E  P  E  E  E  I  D  E  N  E  E  I  T  T  T  F  F
                           470                 490
            GAACAGTCAGTGATTTGGGTTCCCGCAGAAAAGCCTATTGAAAACAGAGA
             E  Q  S  V  I  W  V  P  A  E  K  P  I  E  N  R  D
                  510                 530                 550
            CTTCCTGAAAAATTCTAAAATTCTGGAGATTTGCGATAATGTGACCATGT
              F  L  K  N  S  K  I  L  E  I  C  D  N  V  T  M
                           570                 590
            ACTGGATCAATCCCACTCTAATAGCAGTTTCAGAATTACAGGACTTTGAG
            Y  W  I  N  P  T  L  I  A  V  S  E  L  Q  D  F  E
                  610                 630                 650
            GAGGACGGTGAAGATCTTCACTTTCCTACCAGTGAAAAAAAGGGGATTGA
             E  D  G  E  D  L  H  F  P  T  S  E  K  K  G  I  D
                           670                 690
            CCAGAATGAGCAATGGGTGGTCCCGCAAGTGAAGGTGGAGAAGACCCGCC
              Q  N  E  Q  W  V  V  P  Q  V  K  V  E  K  T  R
                  710                 730                 750
            ACACCAGACAAGCAAGCGAGGAAGACCTTCCTATAAATGACTATACTGAA
            H  T  R  Q  A  S  E  E  D  L  P  I  N  D  Y  T  E
                           770                 790
            AATGGAATTGAATTTGACCCAATGCTGGATGAGAGAGGTTACTGTTGTAT
             N  G  I  E  F  D  P  M  L  D  E  R  G  Y  C  C  I
```

Figure 1 (continued)

```
                 810                 830                 850
SEQ ID NO:3 TTACTGTCGTCGAGGCAACCGTTACTGCCGCCGTGTCTGTGAACCTTTAC
SEQ ID NO:4   Y  C  R  R  G  N  R  Y  C  R  R  V  C  E  P  L
                                870                 890
            TAGGCTACTACCCATACCCCTACTGCTACCAAGGAGGTCGAGTCATCTGT
            L  G  Y  Y  P  Y  P  Y  C  Y  Q  G  G  R  V  I  C
                    910                 930                 950
            GTGTCATCATGCCTTGCAACTGGTGGGTGGCCCGCATGCTTGGGAGAGTC
             R  V  I  M  P  C  N  W  W  V  A  R  M  L  G  R  V
```

Figure 2 Human ChMIrp

```
                    10                  30                  50
SEQ ID NO:1 GCAACTCCACCTCAGCAGTGGTCTCTCAGTCCTCTCAAAGCAAGGAAAGA
                              70                  90
            GTACTGTGTGCTGAGAGACCATGGCAAAGAATCCTCCAGAGAATTGTGAA
SEQ ID NO:2                    M  A  K  N  P  P  E  N  C  E
                   110                 130                 150
            GACTGTCACATTCTAAATGCAGAAGCTTTTAAATCCAAGAAAATATGTAA
             D  C  H  I  L  N  A  E  A  F  K  S  K  K  I  C  K
                              170                 190
            ATCACTTAAGATTTGTGGACTGGTGTTTGGTATCCTGGCCCTAACTCTAA
              S  L  K  I  C  G  L  V  F  G  I  L  A  L  T  L
                   210                 230                 250
            TTGTCCTGTTTTGGGGGAGCAAGCACTTCTGGCCGGAGGTACCCAAAAAA
            I  V  L  F  T  G  S  K  H  F  T  P  E  V  P  K  K
                              270                 290
            GCCTATGACATGGAGCACACTTTCTACAGCAATGGAGAGAAGAAGAAGAT
             A  Y  D  M  E  H  T  F  Y  S  N  G  E  K  K  K  I
                   310                 330                 350
            TTACATGGAAATTGATCCTGTGACCAGAACTGAAATATTCAGAAGCGGAA
              Y  M  E  I  D  P  V  T  R  T  E  I  F  R  S  G
                              370                 390
            ATGGCACTGATGAAACATTGGAAGTGCGCGACTTTAAAAACGGATACACT
            N  G  T  D  E  T  L  E  V  R  D  F  K  N  G  Y  T
                   410                 430                 450
            GGCATCTACTTCGTGGGTCTTCAAAAATGTTTTATCAAAACTCAGATTAA
             G  I  Y  F  V  G  L  Q  K  C  F  I  K  T  Q  I  K
                              470                 490
            AGTGATTCCTGAATTTTCTGAACCAGAAGAGGAAATAGATGAGAATGAAG
              V  I  P  E  F  S  E  P  E  E  E  I  D  E  N  E
                   510                 530                 550
            AAATTACCACAACTTTCTTTGAACAGTCAGTGATTTGGGTCCCAGCAGAA
            E  I  T  T  T  F  F  E  Q  S  V  I  W  V  P  A  E
                              570                 590
            AAGCCTATTGAAAACCGAGATTTTCTTAAAAATTCCAAAATTCTGGAGAT
             K  P  I  E  N  R  D  F  L  K  N  S  K  I  L  E  I
                   610                 630                 650
            TTGTGATAACGTGACCATGTATTGGATCAATCCCACTCTAATATCAGTTT
              C  D  N  V  T  M  Y  W  I  N  P  T  L  I  S  V
                              670                 690
            CTGAGTTACAAGACTTTGAGGAGGAGGGAGAAGATCTTCACTTTCCTGCC
            S  E  L  Q  D  F  E  E  E  G  E  D  L  H  F  P  A
                   710                 730                 750
            AACGAAAAAAAAAGGGATTGAACAAAATGAACAGTGGGTGGTCCCTCAAGT
             N  E  K  K  G  I  E  Q  N  E  Q  W  V  V  P  Q  V
                              770                 790
            GAAAGTAGAGAAGACCCGTCACGCCAGACAAGCAAGTGAGGAAGAACTTC
              K  V  E  K  T  R  H  A  R  Q  A  S  E  E  E  L
                   810                 830                 850
            CAATAAATGACTATACTGAAAATGGAATAGAATTTGATCCCATGCTGGAT
            P  I  N  D  Y  T  E  N  G  I  E  F  D  P  M  L  D
```

Figure 2 (continued)

```
                                870                 890
SEQ ID NO:1 GAGAGAGGTTATTGTTGTATTTACTGCCGTCGAGGCAACCGCTATTGCCG
SEQ ID NO:2  E  R  G  Y  C  C  I  Y  C  R  R  G  N  R  Y  C  R
                   910                 930                 950
            CCGCGTCTGTGAACCTTTACTAGGCTACTACCCATATCCATACTGCTACC
              R  V  C  E  P  L  L  G  Y  Y  P  Y  P  Y  C  Y
                         970                 990
            AAGGAGGACGAGTCATCTGTCGTGTCATCATGCCTTGTAACTGGTGGGTG
            Q  G  G  R  V  I  C  R  V  I  M  P  C  N  W  W  V
                  1010                1030                1050
            GCCCGCATGCTGGGGAGGGTCTAATAGGAGGTTTGAGCTCAAATGCTTAA
             A  R  M  L  G  R  V
                         1070                1090
            ACTGCTGGCAACATATAATAAATGCATGCTATTCAATGAATTTCTGCCTA
                  1110                1130                1150
            TGAGGCATCTGGCCCCTGGTAGCCAGCTCTCCAGAATTACTTGTAGGTAA
                         1170                1190
            TTCCTCTCTTCATGTTCTAATAAACTTCTACATTATCACCAAAAAAAAAA
              1206
            AAAAAA
```

Figure 3 Homology of Human and Mouse ChMIrp

```
Human (SEQ IN NO:2)1 MAKNPPENCEDCHILNAEAFKSKKICKSLKICGLVFGILALTLIVLFWGS 50
                     |||||||||| |||||||| ||||||||||||||||||||||||||||
Mouse (SEQ ID NO:4)1 MAKNPPENCEGCHILNAEALKSKKICKSLKICGLVFGILALTLIVLFWGS 50

                  51 KHFWPEVPKKAYDMEHTFYSNGEKKKIYMEIDPVTRTEIFRSGNGTDETL 100
                     ||||||| || ||||||||||||||||||||:||||||||||||||||||
                  51 KHFWPEVSKKTYDMEHTFYSNGEKKKIYMEIDPITRTEIFRSGNGTDETL 100

                 101 EVHDFKNGYTGIYFVGLQKCFIKTQIKVIPEFSEPEEEIDENEEITTTFF 150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
                 101 EVHDFKNGYTGIYFVGLQKCFIKTQIKVIPEFSEPEEEIDENEEITTTFF 150

                 151 EQSVIWVPAEKPIENRDFLKNSKILEICDNVTMYWINPTLISVSELQDFE 200
                     |||||||||||||||||||||||||||||||||||.||||||||
                 151 EQSVIWVPAEKPIENRDFLKNSKILEICDNVTMYWINPTLIAVSELQDFE 200

                 201 EEGEDLHFPANEKKGIEQNEQWVVPQVKVEKTRHARQASEEELPINDYTE 250
                     |:||||||| .|||||:|||||||||||||||| ||||||:||||||||
                 201 EDGEDLHFPTSEKKGIDQNEQWVVPQVKVEKTRHTRQASEEDLPINDYTE 250

                 251 NGIEFDPMLDERGYCCIYCRRGNRYCRRVCEPLLGYYPYPYCYQGGRVIC 300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
                 251 NGIEFDPMLDERGYCCIYCRRGNRYCRRVCEPLLGYYPYPYCYQGGRVIC 300

                 301 RVIMPCNWWVARMLGRV 317
                     |||||||||||||||||
                 301 RVIMPCNWWVARMLGRV 317
```

Figure 4 Homology of Multiple Gene Family Members

```
SEQ ID NO:      1                                                  50
 2    HuChMIrp  ~~~~~~~~~~ --MAKNPPEN CEDCHILNAE AFKSKKICKS LKICGLVFGI
 4    MuChMIrp  ~~~~~~~~~~ --MAKNPPEN CEGCHILNAE ALKSKKICKS LKICGLVFGI
 5      MuChMI  MTENSDKVPI TMVGPEDVEF CSPPAYTTVT VKPSGSPTRL LKVGAVVLIS
 6     RatChMI  MTENSDKVPI TMVGPEDVEF CSPPAYATVT VKPSGSPTRL LKVGAVVLIS
 7  BovineChMI  MTENSDKVPI ALVGPDDVEF CSPPAYAAVT VKPS.SPARL LKVGAVVLIS
 8      HuChMI  MTENSDKVPI ALVGPDDVEF CSPPAYATLT VKPS.SPARL LKVGAVVLIS
 9     bitChMI  MTENSDKVPI ALVGPDDVEF CGPPAYATVT VKPS.GPARL LKVGAVVLIS

SEQ ID NO:      51                                                 100
 2    HuChMIrp  LALTLIVLFW GSKHFWPEVP KKAYDMEHTF YSNGEKKKIY MEIDPVTRTE
 4    MuChMIrp  LALTLIVLFW GSKHFWPEVS KKTYDMEHTF YSNGEKKKIY MEIDPITRTE
 5      MuChMI  GAVLLLFGAI GAFYFWKGND NHIYNVHYSM SINGKLQDGS MEIDAVNNLE
 6     RatChMI  GAVLLLFGAI GAFYFWKGND NHIYNVHYTM SINGRLQDAS MEIDAANNLE
 7  BovineChMI  GAVLLLLGAI GAFYFWKGSD NHIYNVHYTM SINGKLQDGS MEIDAGNNLE
 8      HuChMI  GAVLLLFGAI GAFYFWKGSD SHIYNVHYTM SINGKLQDGS MEIDAGNNLE
 9  RabbitChMI  GAVLLLFGAI GAFYLWKGSD NHIYNVHYTM SINGKLQDGS MEIDARNNLE

SEQ ID NO:      101                                                150
 2    HuChMIrp  IFRSGNGTDE TLEVHDFKNG YTGIYFVGLQ KCFIKTQIKV .IPEFSEPEE
 4    MuChMIrp  IFRSGNGTDE TLEVHDFKNG YTGIYFVGLQ KCFIKTQIKV .IPEFSEPEE
 5      MuChMI  TFKMGSGAKE AIEVNDFKNG ITGIRFAGGE KCYIKAQVKA RIPEVGTVTK
 6     RatChMI  TFKMGSGAEE AIEVNDFQNG ITGIRFAGGE KCYIKAQVKA RIPEVSTGTK
 7  BovineChMI  TFKMGSGAEE AVEVNDFQNG ITGIRFAGGE KCYIKAQVKA RIPEVGTMTK
 8      HuChMI  TFKMGSGAEE AIAVNDFQNG ITGIRFAGGE KCYIKAQVKA RIPEVGAVTK
 9  RabbiCchMI  TFKMGSGAEE AIEVNDFQNG ITGIRFAGGE KCYIKAQVKA RVPEVGTVTQ

SEQ ID NO:      151                                                200
 2    HuChMIrp  EID....ENE EITTTFFEQS VIWVPAEKPI ENRDFLKNSK ILEICDNVTM
 4    MuChMIrp  EID....ENE EITTTFFEQS VIWVPAEKPI ENRDFLKNSK ILEICDNVTM
 5      MuChMI  QSI.SELEGK IMPVNYEENS LIWVAVDQPV KDSSFL.SSK ILELCGDLPI
 6     RatChMI  QSI.SELEGK IMPVKYEENS LIWVAVDQPV KDNSFL.SSK ILEFCGDLPI
 7  BovineChMI  QSISSELEGK IMPVKYEENS LIWVAGDQPV KDNSFL.SSK VLELCGDLPI
 8      HuChMI  QSISSKLEGK IMPVKYEENS LIWVAVDQPV KDNSFL.SSK VLELCGDLPI
 9  RabbitChMI  QSISSELEGK IMPVKHEEEA LVWVAVGQPV QDNSFL.SAR VLELCGDLPI

SEQ ID NO:      201                                                250
 2    HuChMIrp  YWINPTLISV SELQDFEEEG EDLHFPANEK KGIEQNEQWV VPQVKVEKTR
 4    MuChMIrp  YWINPTLIAV SELQDFEEDG EDLHFPTSEK KGIDQNEQWV VPQVKVEKTR
 5      MuChMI  FWLKPMYPKE IQRERREVVR NS.APSTTRR PHSEPRGNAG PGRLSNGTRP
 6     RatChMI  FWLKPMYPKE IPRERREVVR SS.APSTTRR PHSEPRGNAG PGRLSNRTRP
 7  BovineChMI  FWLKPTYPKE IQRERRELVR KIVTTTTTRR LRSGPQGTPA PGRPNNGTRP
 8      HuChMI  FWLKPTYPKE IQRERREVVR KIV.PTTTKR PHSGPRSNPG AGRLNNETRP
 9  RabbitChMI  FWLKPTYPKE IQRERREVVR KTV.PTTTKR PHSGPRGNPG PARMRNDSRP

SEQ ID NO:      251                                                300
 2    HuChMIrp  HARQASEEEL PINDYTE... NGIEFDPMLD ERGYCCIYCR RGNRYCRRVC
 4    MuChMIrp  HTRQASEEDL PINDYTE... NGIEFDPMLD ERGYCCIYCR RGNRYCRRVC
 5      MuChMI  NVQDDAEPFN PDNPYHQQEG ESMTFDPRLD HEGICCIECR RSYTHCQKIC
 6     RatChMI  SVQDDEEPFN PDNPYHQQEG ESMTFDPRLD HEGICCIECR RSYTHCQKIC
 7  BovineChMI  SVQEDAEPFN PDNPYHQQEG ESMTFDPRLD HEGICCIECR RSYTHCQKIC
 8      HuChMI  SVQEDSQAFN PDNPYHQQEG ESMTFDPRLD HEGICCIECR RSYTHCQKIC
 9  RabbitChMI  SVQEDSEPFN PDNPYH.QEG ESMTFDPRLD HEGICCIECR RSYTHCQKIC
```

Figure 4 (continued)

SEQ ID NO:		301				337
2	HuChMIrp	EPLLGYYPYP	YCYQGGRVIC	RVIMPCNWWV	ARMLGRV	
4	MuChMIrp	EPLLGYYPYP	YCYQGGRVIC	RVIMPCNWWV	ARMLGRV	
5	MuChMI	EPLGGYYPWP	YNYQGCRSAC	RVVMPCSWWV	ARILGMV	
6	RatChMI	EPLGGYYPWP	YNYQGCRSAC	RVVMPCSWWV	ARILGMV	
7	BovineChMI	EPLGGYHPWP	YNYQGCRSAC	RVIMPCSWWV	ARILGMV	
8	HuChMI	EPLGGYYPWP	YNYQGCRSAC	RVIMPCSWWV	ARILGMV	
9	RabbitChMI	EPLGGYNPWP	YNYQGCRSAC	RVVMPCSWWV	ARILGMV	

Figure 5 (SEQ ID NO: 10)

```
   1  GTGCTGTTCC TTTGGCCTGG AATAATCTTT TCCCCACTCA TTACCAGCTA
  51  ACTTCTACTC ATTATTCAGG TCTTAGCTTA AGTGACAACT CCTCAGCATG
 101  GATTTCCCTC ACCCATCCCA CCCAATCCAA ATTAGGTGCT TTAAAGCACC
 151  TGCACTTTTC CTTCTTAGCA TATATCATCA TTTATGTACT TTTATTTACA
 201  GGATTTTTTT CCTTTTAGTT TAATGTCTGT CTACCCAACA AGACTCTGAG
 251  CTCTATAAAG GAGGAACTGT GTTTGCTTGT CTACTACTGA ATATATATCT
 301  AGCTCCTATA ACAATGCCTG CCACATAGGG GACACTGAAT AAATACATGC
 351  TAATCAATGA ACAGTCTATT GAAATGCCAT GTCCTTTGTG CCATTTTTCC
 401  TGCTTTTCTG AATCAGAATT AATCACTCTT CCCTCTATGT TTCCATGGAA
 451  TATTGCTTTG ACTCTTGGTT TAATTCAATC CTTTTAACAA GAAAGGCATG
 501  TTTTAAAAGA TCTCAAAGTC TATTCAGGAG AGGAACACAT AAAAAAGAAT
 551  TATAATACAG CATACAAAGT ATAACGGTTG AAGTATGAGC AAGATACAAA
 601  AGTAGAGCAA GGGAAAGGGT AATTAACTTA ACGAAGTGGA AGAGATGAGA
 651  CCAGGAAAGG CTTTCAGGGG GTGATAACAT CTAGGCAGCA TTTTGAAAGC
 701  CAAATGGGAA TTTTCCAGGT AAACAAGGGA AGGCAGAAAA CTCCAGGCAA
 751  TGAATATAGC ATGGGCAAAT GTCCTACAGT GTGAAGAAGT AGGGCATGTT
 801  CATTAATAGG GTGTACAGTG TGAGGCAGGC ATCAGAGGCA TAGTGTGAAT
 851  AACATCACTG AGGGGAACCA TCTGACCCAA TAAACATAAG TTTCTCCATG
 901  AGACTGCTCT TGTGGCTTTT CTTTTTTCTC TCTTTAAGCT TGACCAAAGA
 951  ATTGTAGAAC ATATAGAATC GTGACATATT TCATAATTTA TGTAATCAGT
1001  TTAATGCACA ATCATGTATC CAACAATTAG AAAATTATTT TCCTTGAGAC
1051  ATTAGGATGT GGAGGTAGTC ATATGTACAT GCCCCATTTT CCTGCCCAGC
1101  TGCGACCTCC CAGTTGGCTC CATCGCTGTA GAAAAAAATT GTCCAAAATC
1151  AAGACTAGAG AGTTGGACAA AGTCGAGATC ATTAAATGCC TACTATGTCA
1201  CAATAAGTGG CTTGGATTTT ATAGACAATG AATAGATACT AAAGGATTTT
1251  AAGCAAGGGA GAGATTATCC TCAAATTTGA GTCTTAGAAA GATCACCTTG
1301  AAGACAATGT ATGAGAGACA AGGAAAACAG AATTCTATTG CAATAGTAAA
1351  AGTGACTGAG AACCTCAACT TAAACAGTGG TGATTGACAA TGGAGAGAAT
1401  AGGATTGCAG AGCTAATTCA GAGGCAGAAT TAACTGGACT TAATAATTGA
1451  TTAGCCATGG GGATGAACAA GACAGGAAGA AATCAAGGAT CATCTGGAAG
```

Figure 5 (continued) (SEQ ID NO: 10)

1501 TTTCTATTCT AGAGACTGAT TATTGTGTCA TTTAGTGAGA GAGGTAATAC

1551 CAAAAAAAAA GTGCATGTTT GAAAGGAGAG ATAACTAATT CAGTTCGGGA

1601 TATATGGAAT TTGAGGTGTC TGTTGGGTAT TGTAGATGTC TAGCAGGTTA

1651 GATGGACATA TGGATGTGGA GCTCCAAGAG AACAGCCTGG GTAAAATATA

1701 TGGATGTGAA AGTTCTCAAA AAGGATGACA AACTAAACTG GGGTGTTAGC

1751 AGTGAGCACA ATAGGAAAGA ACAGCTTTCA AAGATATAAA AATTTAGAAT

1801 AGATAGGATC TAGTGATTGA TTGGAATAGA TAGGATCTAG TGATTGATTG

1851 GATAAAGGGA ATGTGAAAAA GAGACAAAAC CTCAGATTTT TGGTTTGAAA

1901 GACAAACCTC ACATTTTTGG TTTGAAAGAC AAACCTCAGA TTTTGGTTTG

1951 AAAGACTGGG TGCGTGACAA TTCCACTATT ATAAGCTAAG TTATGTTCCC

2001 TCAAAATTCA TATGTTGATG CCCTAGCCCC CAGTATTTCT ATATTTGACT

2051 GTATTTGGTG ATAGGGTTTT TAAAGAGGTA ATTAAGGTTA AGTGAAGTCA

2101 TATGAATGGG ACCTTATCCA ATATAACTGA TGTTCTTATA ATGAAAGGAG

2151 ATTAGGATAC AGGCATACAT AGAGGGAGAA TGTCATGTGA ATATGGAGAT

2201 GGTCATCTAC AAGCCAAGGA GAGAGGACTC AGAAGAAGCC AACCCTGATA

2251 ACACCTTGAT CTTGGACTTC TAGCCTCCAG AGCAGTAAGA ATATAAATTT

2301 CTGCCATTTA AGCTACTCAG TCTGTGGTAT TTTGTTATGG CAGCCCTAGC

2351 AAACTAATAT AGCCACCAAC TAAGGTTGGG AATATGAGGG GAAAAAGTAG

2401 GTTAGGAGAA GGTGTTGAAA GAAGACTTTG ATATAGAACA TGTTGAGTTT

2451 GAGATTCCTG AGGTATATCC AGGTGGAGAT GTTCAAATGG CCTTGAAATT

2501 AAAAGATATA ACTGGGTTAC AGGTACAGAT GTGGAAATCA CTGATATGTA

2551 GGTGATATTT TCAACTGTAG GAATGAATGA TATCATCTAC GAAGTCTGTG

2601 TAAAGTGAGA ATGATCATCA AGAAGAGAAC CATAAGGCAT GTGAACATTA

2651 AAAAGCAGAC TGAGGAAGAG GACCCAAGAA AGGAGAAAAA TCAGAGACAG

2701 AAGATGAAGA GAGAGAGCTT AGAATTAGTC TTTATTTTAT CCTGTTTTCA

2751 TTATAATTAG TTAAGTACAT ATTGATAGCA CCATTTAAAG AGCTTTAACT

2801 TAGTGAAGGA GGCAATATCT TATTCATTTC TGAATCTCAA ACACCAAGCA

2851 CAGTGCCTGA CATTTAATAT CCCTGCAATA AATATTTGTT GAACCAAATG

2901 GAATCGAAAT TAGGACTTCA TTCCCTTCTT TGTTTTGTTT GCTTATTTCT

2951 GGAGTGTGCA GCATGTATTG CCACTGATAT TTTAAGTTCA TAGTGAAAAA

Figure 5 (continued) (SEQ ID NO: 10)

3001 TCATTTCAAT GTGTTCTAGG CTTGGTCCCC TTCCAACCCA ACCCCCAGGA

3051 CAATAGTTTT CATTATAACC TTCTGTCTTT GCCATTCTGG ATGGAATTCT

3101 GTGCACAGAA GTTATATACA TATATGGGTA TATCTATGTA ACAAATCGCA

3151 GCACAGGAGT CCCCTGGGCT CCCTCAGGCT CTGGTATGAC ATATTTGAGC

3201 CATATAAATT CAGCTTCTCC TCTGGCATCT GTTAGCCGAC TCACTTGCAA

3251 CTCCACCTCA GCAGTGGTCT CTCAGTCCTC TCAAAGCAAG GAAAGAGTAC

3301 TGTGTGCTGA GAGACCATGG CAAAGAATCC TCCAGAGAAT TGTGAAGACT

3351 GTCACATTCT AAATGTAAGT TGATTCATAT TTTTTCCCTT TTGAGCAGAA

3401 GCATGGTTTC ACAGATTTAT CATATTGCAA AGTGAACATT AGAAAGTGGA

3451 ATCAAAGGTG ACTTTGAATT ATGGCTTGCT TATTGAGATT TATTGTGTAT

3501 TGTTTTATTC TCTCTGTTCT TTGGTTAGGC AGAAGCTTTT AAATCCAAGA

3551 AAATATGTAA ATCACTTAAG ATTTGTGGAC TGGTGTTTGG TATCCTGGCC

3601 CTAACTCTAA TTGTCCTGTT TTGGGGGAGC AAGCACTTCT GGCCGGAGGT

3651 ACCCAAAAAA GTAAGTAAAT ACACATCATA ATCTGATGCT TCTGTTCTGA

3701 GTTTGATTGA ATTTAATTAG TAGGCATATA AATTATTCTG AAAATGAAAA

3751 TCATTAAGTC TCTTTTGTGT TTATTTTCTG GTGGTATGAA AAACTGAATC

3801 AAGATTTTTC TCTCCTTCAT TTGCCTAATT ATTTTGGTAA GGGAAAAAAA

3851 TTTATCTTCT ATATTTTGGT TAGTAATGAT AGGGTGTAGA ATTACCCATT

3901 CTTCTATAAT GCTTATATTT AAGGAGTTAA GACTCCTCCT AGGAGCAAAG

3951 TTTTTACCTT AAGAAAAAAT ACTTTTAATC TTTCTTTCAA TTATCTGGGC

4001 AGAAACTGAT GGAAAAGTAA TTAGACCACC ACATAAGGAG AAACCACCCA

4051 AAGTGGCAGC AAAGGATTTT CAGGGATTAT GTTTCTGACT TAATGCATCT

4101 GCAGTGAGAA CAGGCAGCAA GCTGCTTTTC AGGAGGAGCA ATTCTCATAC

4151 CTGACTTCAC ATTTATTTAT GACTTTAAAA AGAAACTAGA GCTTTTCAGC

4201 TCCTTGCCAG CTGCAATCAT CACTGTCCCA CTGAGATAGC AATGCCACAA

4251 GGTAGAAGCA TGCGACACCA AGTCATTAAT CACAGATAAC TCAGAAAGCT

4301 TCTGATACTG GGGTCATTTG GTGAAGCTTG GCTAGCCACT GTAGTTTATA

4351 TTTCTGTACC TTAGTGTTAT TGTTATTCCT GGGCTCATAG GCAATGGGCA

4401 AGTTTAGATA TAAAATAAAT AAATAAAACC ATAGTCCCAA ATAAAACCAT

4451 AGTCCCTTCA TGCTCACAAT TTCCATTTGA AAGATAATCA ACAAATATAC

Figure 5 (continued) (SEQ ID NO: 10)

```
4501  AAACTACAAG TTGTGAAGAT AGTTTTCTGG CTTTGCCATG TTTTTCTGTT
4551  GGTGTGTGGA TTGCTGCTGA AGTTATTGTT ATTACTGTTG TTAATTGGTT
4601  GTCATAAGAT GCTCATTTCT ATTTGGTAAC AGATGTAGGG AGAAAAAAGC
4651  ACAAATCCAG GAAGGAGGAT GATACAAGAA AATGGGTTGG CATGGACCGT
4701  TAGGAGCCAA AGGAGCAATA GGAGCAGAGA GAAAAAAATA TGGGTCAGTG
4751  GGGAGAGAGA GGCAGAGCAG GAAAAGCAGA AAATAAGCTG GCTAAGTGTC
4801  TGAAAGATAA CAGAGGTATA AAGCATAACC ACCCTTGGGT TCTCCACCTC
4851  TGCTGTCATA CCCACTTTCA TTTAGTCCCA AAGGATATCT TATCAGGTTT
4901  AAGTAGGAGG GAGAGCAAGC CAGCCTACTT TGTTTGCCTT TTTATTTATA
4951  CTTTGTTCCA GAAAAGATTT CAGGCATCAC CCAATGCCTA TAGTTAGCCC
5001  TATCCCCTAC TCCTCCTCCC CTTACTGCTC CTTGCTTCTG AACTACTTAC
5051  CGTTCAAATG TCCTTCTAGA GAAGCCACCT GTGTTCAGGC TGGCAGGGTG
5101  GGTAAAGCAA TCGAGATCAG GAAGGAAACA GCAAATTAAT AAGAGAGCCA
5151  AAATGCATTC CAATACAATT CTGGTACCAC TTAGCTAATT CCCAAGGTGA
5201  TATGTCACCC CACAAGTTGA ACCTTGGATC CTCTTGAAAA TTATCCATGA
5251  GTTAAAGGGC CATAAATCCT TTGTTCTCCA CAGCCATAAT GTACTTAACT
5301  GGGCATGGTG GATAGCTAGT CCTTTTGACC ACCCTTAGAT GCTTGCCTAA
5351  ATAAGGGGCC AAATGGCCTT TCTTGTTTTA TTGTTGTCAT TGGTACTGCA
5401  AATGCTATTC CAAAAGAGTA TATGTGTTTT TATAAATTTA CAGTGCTATT
5451  TGCCTCAACG CAGCTATAAA CGACTAGGTG GAGACCTTGC CTTTTTATTG
5501  CAAAGTGTCA AACTCACTTG CAGCTTCATG GTCTCCTCTC TTTCCCTATT
5551  ATATCATCTT TACATCTCAT TCTAGTTCTG TTTTTAATGA TTATAAGGAG
5601  CAATCCTATG ATACCTCAAA CCAAAATAAA ACTTGGTCAG TGCATGCACA
5651  GAAGAGGTGG GATATGCTGC TGAATAAAGT TTATCAAAAA TCAAACCTCT
5701  TGTGAAGTAA TCCATTTTGG GATGTCATTT TCAGCAAAGG GACTCTCAGA
5751  AATGAACAAA GCTATTATGA AGTAATCCAG ATAAAGACAA CAGGTGTCCC
5801  CTAGGGCAAG TTTCTCTACC ATGAACAAAT ATTGTTGGTT TTCCATTCCA
5851  GTGGGGAGTT ATGATTCATT GTCTTTGATA TTTAAGTCAT GGCCTTCATA
5901  AGACCAGGTT TGGTTCATTG AAAACTAAGA AATCAGCTGG AAATGAGGAT
5951  GCTTTACTTA GTCCACTACT ACCTTCAGGT CCCAGACAAA GATCTCAAGA
```

Figure 5 (continued) (SEQ ID NO: 10)

6001 GACCATCTAT GTTTCCAATG GGTACGCCTA TCAACAAGCC TGCCTTAGAG

6051 CAGACTTAAG AATTCTGGGA AGAGACTAGC AGTCAGATTA GTCTTGGAAG

6101 AGTTGCATAG GAAAGGGACA CAGGCATCTT ACATAAATTA AAGACTCTCC

6151 CCTTAATGGT ATGGCCTAGC AAGAAACAAC AGCAGATTAA AATCCAGGCT

6201 CTTTGGGCTT TCCTTAACAA ATCTATGAGA CTTGTTGGGA CTCTGTGGCC

6251 CTTGGTAGAG AGAGCTGTGC ATGCAACTGT GCTTATTGGT GGCAGTGATT

6301 TGTTTGCAGT CCTGGGATTT CCTCCAAGCT CTTAGTGTTT GCTCTCCTAG

6351 CCTCCACCAC GCCTCCTTTT TATTTTGTCT TTCAACATCT GTGTATAATA

6401 AACAGTATTT ATTTAACATC CCACATCTGG TGTTGACATT TCCATGCCAA

6451 TAAGTCCTTC TCCTCACTCC CGCTTTGGTG TGTACAAAAG ATCCAGATTC

6501 CTGTATCCAA ATTGAGAACA AGCCATGCTT GACCTCCTCC CCTCACATAC

6551 CTCAGGTCTG CGTGCACACA GGTGTGCACA CGCACAGACT CACACACACA

6601 CAGACACACA CACACACACA CACAGACCAG AGAGAGAATG AAATCAAAAT

6651 AAAGTGAAAC AAGGACGACT AACTCTGGGT GAGCTGCTGC ATTTTCTTAC

6701 AGGTCCCAGG CCCCTGACCC ACAGCCACAT ACAGGAGCCA CCTGAGCTAA

6751 ATGAAAGCCA GGTCCATCCA GCTCTGAATT CAGCTTTCTT AAACAAGCAG

6801 AAGAGACCTA TTCAGTCCCA GCAATTTTGA GAAAGCATCG TAATGAGGCT

6851 CCCCTTCTAG GCAGTGACTG CCCTGGGATC ATTGAAGCAG AATCACCAAA

6901 GAAACATTCT TGCTTAAAGC CCCCTGTATT ACACATTTTA CTGGAAACAG

6951 TCAGAAGCCA ACTCCTCAGC TAATTAATCC AACATGAAGC AAACAAATCA

7001 GAAAAATCCT AGCCCTTAAA CCAACAGTTT AGAAACTTAG CACAGACCTT

7051 GTAGCCCATC CCTGCCAATC CTTTGTGGAG CGAGGCAGAA GAATCATACA

7101 TTTAATAAAT GAAAGAGTGT CCTTACTGCC TTACATAGGA GCTGCCTGGT

7151 GACTTTAGGG TTGTTTGAAA GCCATCAGGA CCACAGTAAT CAATAGATTC

7201 GTGAGGATGT TCCTTCAACT TCTATATATT ACCAGAGCAA TGAGGAGTGG

7251 TAGGGATAAG CGGGCACTTT TCTGCCCTCC TGCCACACCT CAGTTGCAGG

7301 TGATCACTAA AACAGTCTCC CCCAGCGTTG TGAGTCCACT ATCATTCTAG

7351 GTTGAAAATC CTTTGCTCAG AAGAAGCACT TTCTTGACGA GTCTGATTGC

7401 AACATTTTGT GCGTGCCTAT TTCTTGGCAC GCACAATTGG GCATTAGCCC

7451 AATATTAACT TTTTTCGCTC CTCTCTAGGC ACACAAGGAA GACAACTTTT

Figure 5 (continued) (SEQ ID NO: 10)

7501 TTCTCCAGGT CCCCTGGGCT TTCCTACACC ACAGTCCGGT GGGTGAAATG

7551 TGGGTCTAGA GAAACATTAA TATACCCTCA AGGAAAAAGT TGAAGGAGTA

7601 GCCGGAATAT TTTTGGCAGG GAGACTGAGC TTTACCAGAT GGTGCCATGT

7651 TGGCATTTTC AAAAAAAAAA AAAAAAAAAA GACCTCTCTT CATTTTCCTC

7701 TTAAAGCATT GTGAGACCCA GGTCCACCCT TGTACAACTC TCACTGGGCC

7751 AGCTCTTATT CTGGGCTACT CAAGGGCAGC TCCCAGCAAT TATGTGGAGC

7801 CAGGCTGTGG CTGTGATGGG CATGTTCTTC TTAGTCAGTC TCTTCATTCT

7851 CCTAAGGCAG CATTTGCTGA ACAACTTATA TGACACCAGG ATGAAGAGCC

7901 TGGGCCTTCA AAGACAGCTG TTCTGGAGTG GTTGAGAGTG GAATTGGGGC

7951 TGTGGAGGAA GTAGCCTGGA TGATCAACAG ATAGTGGCAG TAGGCGTAGT

8001 CCAGTGATCA ACCTGACAGG TTGATCTGAG CAAATCTCAG CAGACTTTGC

8051 TAACTATCTA ACATTACAGA TGAAAAGGGG TAAAGGATTA AAATAGGAAC

8101 TGAGTCCCTA GAGACACTCA GTGATTTGCC CAAAATCACA CAGCTTGTAA

8151 GTGGTGGGGT TAAGGTTAGA ACTAGAGGCT AGTCTGAAGA TTCTTTCCAC

8201 TATTCTTCAG CTGCATTTAA GAATCAGGGA TGGGGATAAT GGGTAGGAGA

8251 AAAACTTACT GGTCAGCTCA CGTCTGTAGA TAAAAGAAGT AGCTTTTTCT

8301 TGTTGAAAAA GAATGTCTGT GCTGCTTATG CCAATATTTT GCATATTGCC

8351 AAATGGTGAT TGAAGATGCA GAACTCTAAC TTTGCATAAG GAACAATCTG

8401 TCTGCTCTGT GTTGGTAAAG AAATACCCAG AGGCCCTTTA ATTCCCCGCG

8451 AGTTAGGTTA ATAATAGATA TGGTAGAGTG CATTCTCTTC CCTTCCTTCC

8501 AGCCTAACAC CTGCTAGCCA GCTATTCTGA GCATGTATTT CACCTTTGGA

8551 GAGTGTACCT CATAACAGAC AGTAAGGGAT GAAGATGGAG TGAAGAGCAG

8601 CTAGATTAAT CACAGGCTTC AAAAAACAAA CCTAAGCCAA ATGTCCTCAA

8651 ATCCCTGTCA GGACAGGCTA AAGCTTCTTG CCCCTAAAAC AGATCCATCA

8701 CACTGAGGTC TGAGATAGCT TTACCTTTTC TAAATAGCTT TCTTTTCTGA

8751 TCTCAGGGAC TGCCACAGAG GCAAGAAAAC AAGGAAAACT GTGAAAGAGG

8801 GATGTTAGGG CTCATTTGCT GCCTGAATGG AGGATTTAAC TGGACTGAAG

8851 ACAGAATTCC TGGGCAGGGA AGAAAAATGG GTTGTGTCCT CTACATATTT

8901 TGTCAGTGTG GTCTTCTTAC ACTTCTCACT GCCTGGACTT AACCACATAT

8951 CTGTCTCATT TCACTTCTGA TTCCTATTCC ACTCCATTGG CTTCCTTGTT

Figure 5 (continued) (SEQ ID NO: 10)

```
 9001  CCATTCCAGT GTTCTGGTTC TTATAGTTAA GGCCAGCCTT ATGTGTGGGT
 9051  GGAGTAAGAG CCCATTTGAG TGGTGCTGGA AAGATGCCCT TGCCCCCAAA
 9101  TATCTCCCAT TCCTCGTGCC ACTGTGCCTA TTAGATTGGT GCAAAAGTAA
 9151  CTGTGGTTTT AGCCATTGAA AGTAATGGCA AAGTAATAAA GTAATAAAAT
 9201  CCACTATTAG TACTTCATTC CTCTGAGTGA AGTATTGCAC TCAAATTTTT
 9251  AAACAGACAA ATATTCATCA CTAATATCTA AGTGTGAAAT CATGGGTCAG
 9301  ATTGATTTCT TCTTTGTCAT AAGGTACCTG AAGCACCACG AGATTCCTAG
 9351  ATGTCTAGAC AGGAGACAAA GGGACATAAA TGGGAAGAGA AAGGGGAAGG
 9401  GCTATGAGGG TATAGAATGA ACAGAATATG AAAATAGAGC AGAAAAGGAA
 9451  AGCTAGAAGC ACCACAAATT ATCTCTCTCT ATTTTCCAAT TGGGCAGGGG
 9501  TGTACTCCTC ATCATATTCT GTCCTTGCTC TATCCTCCCT GGAATAAAGC
 9551  TGAGTTGGAG GCAGCTGAGA TGAGCTACTT GTAGCCCACT CACACCTCTC
 9601  TTCAGGCCAG TCATCCTGTC TCTGCTTTAG CTTGCTCCTG GAGAGGACAA
 9651  AGCCATTTAT TCTTTCATCT CCAGAAAGAC CTTCAACCCC TGAGCCAGGA
 9701  AGCTTAGTAG AGAGCAAAGT ACTTGTTATG TCAAATGGCA TGAGACTCTA
 9751  CGTCTTCATA CGGGCCAGTG AAGCTTGCAT AGCCATTACC CAAGCAATGT
 9801  GCTCCACAGG GACCCAGGCG AGACAGTGTG TCTGGATTTG CTCATACCCA
 9851  TCTCCAAACA GTCCAAAGTG GGCCACCAGC ATCATCTTCC TAGACGAAGA
 9901  ACAGTCTGTT CCAAAAGGAC ACCTTTGATA GTGGAATTCT GTCTGGGTGA
 9951  ATATATAGCT CCCCACTCCT ACCCCCTACC AAGGATTCAC ACATAAACGA
10001  AGTCATTCTG TTCCCACTCT AGTAACTTAT CATTACTAAC ATAAAGCCAA
10051  CAAGTTGGTT AATGAACACA ATCAGAATGG ACCCCTTCTC CCCTACCAAA
10101  AAAACTCTTT GCAGGAATGT TGAAAATAAT CAAGATCTGC TGAGCTTGGC
10151  AGTTTCCAAT AGTTTTCATC TAGCAAGGTG TCCAAGTAAA GTGCATTGCC
10201  AACAGTGCAG GAAAACTCAT TCCACAGAAG CACACTGAAA TTCTAAATGT
10251  CACCAGTAGG TTCCATAGTC TTCTACCAGT GATTGAAAAT ACAGCCTGTA
10301  TTTTCTCTTT CGCCCAGGTA GGCAGTGTTC CAGCATCCCC TAGTCTCTAA
10351  CCTCATTTTT CCATACAATA ATCCATCAGT GGTGCTACTG ATGAAATAAA
10401  CACCCCACCG CAGCCCCCAA TGGCTAGTGG CATCAACATT TGTCTCAAAA
10451  ATGGATAGTA TTTCTAACAA AATTCCCCCG TTGGGAGAGC TGACAGGACT
10501  GGATCTGAAG TGGATGCCTC TCTTTTCTAC TTAGGTCAGA CAGACTGAGA
```

Figure 5 (continued) (SEQ ID NO: 10)

10551 GCTCAATTTG TTGCCACAGG TAGAAGTTTC TCTTCTCAAG CTGGCTGCAG

10601 AGCTGGAACT GAGGACATAT TTATTGCTAG AAGCAAATAA CAAACCTAAA

10651 GGAATTTGCT ACCTATCACA AACCAGGCCA GACTTCCTGA GTTAACTCTG

10701 CAGTTGCCAG AGAGTTCTAA GTATTGTTTT TCTTTCACTT TTAAAATGGT

10751 TTTTTTAAAA GACCTTGCCA ATACTCAAAA ATGTCTCTTT AGGAAGATAA

10801 GGATGTGCCA TTGGTGGGGA CAACTTATTC AACAAATACT TAGGTTCCAA

10851 CCATGTCCAA AGCACTGTAG GGCTACCAAG AGGTAAATGA CATTGTCTCT

10901 TGTTTTTCTC TTCAGCAGCT ACATTAGACT TGTCACCCTT GAAGTAACAC

10951 TTGCTTCTAC AAATGGTTCC AATTTGTTTT TCTCTAGTTG AGTTGGTGAA

11001 TATTTCACCA GGTTAAACTA ACATTCAGTC AATGGCATCA GCAGAGACTA

11051 AAAAATGACC CAACCTGGTG TAAGTTCTCT ATAAGGCTGC CACTGGGAAC

11101 CGTATCATTA GAAAGCAATG TGGGCTCTGA AGAAGACTTG CTTTTTCAGG

11151 CTTTTAGTTT CCCAGATTTT TTTCATAAGG AAAAATGAAC ATGGAGAAAC

11201 TCGATAAACA AAGGTCTCCT AATGACAGAG ATAAGATAGA CTTCCAAAAG

11251 GCATCATCAG GTCTGGCCTG GTTTGTGACA GATAACAAAT TCCTTTATGC

11301 TTATTTGCTT CTACCTATAA ATATGTCCTA AAACTATAAG TTATGGGCTT

11351 ACTCCTAAAC TGCAGTGCCA TATGGGACTG GTCCTCTTGG TCTTTAGAAA

11401 TACAATTGCC TCTTTTAGAG TGCAGGAGAA AATCTGGGGT TTTGTTTTAG

11451 CCACTCAAGA AAGACCTGAT GAGCAGTCTT TCTTCTTGGG AGGCTCCATG

11501 TTACTTAATG CAGGACACTT GGCTCAGTTG GCTTAATTCC ATATACACAA

11551 AAGAATCCGT AGAACTTGAA AACAACTGAA TATGGCTTGG GAAACAGAAC

11601 CCAGGAATAA ACTGGGTGGG GTTGGTTGCC GTATAAAGAG AAGGCTAAGG

11651 GAAGGCAATA AATGTTTCCT GGAGTTTAAG GGGCAAAAGG AAGCAAACAG

11701 AGCAGGGTAG GGCTCTAATA TTGGCAAGGA GAAAGTCTTC AGATTCTAAA

11751 AGGAATAGAC TCAAGGCTGT TGTGGCAATA TTAAAGCCCT GGATAAATAA

11801 GGAGACAAAA AACACTTAAC CATTTTAAGT GTATTCAAAT CTCTTTCTAA

11851 GCCCTAATGA ATTTCACGGG TGAGGGGCGG GGCAGTAAGG GGGGGGTCGT

11901 GATTACTTAG CACTTGGAAA GGGAAGTGGT AGCTTAAAGA AGCAGCATAG

11951 GTTTTAGGTG ATCCCTCAGC TTAACACAAG GGGAAAATAC TTTATAGGCT

12001 GGTTTGCAAA CTATCATTTG CTGTTTAGTC AAGGCTGCCA AGAAAACTGT

12051 TGGAATTCTT AGTAATTAGC TCAGCAGCTT GGCTTGAATT CAAAATACCA

Figure 5 (continued) (SEQ ID NO: 10)

12101 GCTCTGAAGG GATCCCCATA TCAGCTCCAG CGCCAAAAAG CACCTCACTT

12151 TAGGGACACA CTGAGTATCT TAGAGATTGT TTTCCTCTGT TCTCCCAGGC

12201 CTATGACATG GAGCACACTT TCTACAGCAA TGGAGAGAAG AAGAAGATTT

12251 ACATGGAAAT TGATCCTGTG ACCAGAACTG AAATATTCAG AAGCGGAAAT

12301 GGCACTGATG AAACATTGGA AGTGCACGAC TTTAAAAACG TAAGTTGGAT

12351 GTTTTCCTCC TAAGGCTTTC CACTTAAAAT ATTAGAGCAG TTGAGTCAAG

12401 TTAAAAATAG CCTCCATCTA AAATTTGAAT TCAAATGGAA AGCTAGCCTC

12451 CTCTAGCTTT CCATTTGAAT TCAAATTTTG CTAGTCTGGG ACCCCAACAG

12501 ATCATGGAAC TTACTAATAG TGGCTCTTTT GGGAAGCTTT ATTGTTGTTT

12551 TGTCTGTTTT ATAGGGATAC ACTGGCATCT ACTTCGTGGG TCTTCAAAAA

12601 TGTTTTATCA AAACTCAGAT TAAAGTGATT CCTGAATTTT CTGAACCAGA

12651 AGAGGAAATA GATGAGGTAT GTAAGAAGAA TAATTGTGGT GGCAAAAGAC

12701 ATCATTTATT GGATGCTAGC TATGTGCCAA ACATTGTACT AAATGCTTTA

12751 CTTTTCTTTG ATTCTCCAAC AACCTTATAA ATAGATACTA TTGTTATCAT

12801 CATTTTCAGA TCAGGAAACT AGAACTCAGA GCCAAAAAGT GACTTGCTCA

12851 AGGCCACACA GCTAGTGAGT AGCAGATTGG GGTCTTGAAC TTAGGCCTTT

12901 CAGAAAACTC AAGTGCATAA TTTTAGGTGC AATGCTATAC TCTTATACCC

12951 GGGCACCCAT TAACATTATG TTGATTTTCT CTGAAACCAA AGATTTCTAG

13001 TAAGAGGGAT GAGCCAGAGA AACAAAATCA ATCTTGCCTC AGATAAAGCG

13051 TTTTTAAAAA AGAAAGTTCC AAAGCCACCT GGGGTTCTTG CCACTTTTGT

13101 CACACTTCAG TATTCACCTA TTTTAAAGAT TGTGGTTCAT TCACCTTACC

13151 ATTTTATGCT GGGCTGCTAA TGGAATAATT CACATTTAAC CTCAGTAGGA

13201 AAGGCAAAAG CAAAGAAGGT GAACACTGCA ATGGGATACA AATTCTCTGA

13251 AAACACCCTG CTAGACCAAA GTAAGGCAAC CTAGGGTGCT GGTGAAATCA

13301 CTGAATGGCC CTTTAAAACC TAATGGGAGA TGGATGTTAT TGCCTAAAAT

13351 ACAACGATCT GTAAGGCCCT TGAGGCTCTT TATTGTTCCC TGATATTTTG

13401 TCTATAATTA CAGTATGGTT TTATTAGTGG GAGGAGAAGG GGGACCTCAG

13451 TTAAGGTCAT ATTGTGACAG CATATCCCCA CTATAGCAAG AAAAAACAAA

13501 CATTGGTTTG GTTTGGTTTA GTGCTTAGTT GCCAAATCTA AAGTGGAATT

13551 ATGGCAATGG TAGAATGATC TTTCTTCTTC TTCTTCTTCT TCTTTGTATT

13601 TTTAGTAGAG ACGGGGTTTC ACCATAGCCA GGATGGTCTC AATCTCCTGA

Figure 5 (continued) (SEQ ID NO:10)

13651 CCTCATGATC CGCCCGCCTA GGCCTCTCAA AGTGCTGGGA TTACAGGCAT

13701 CAGCCACCGC GCCCAGCCAT GGCAGAGTGA TCTTTCTATA AACTATAAAA

13751 ACCTCTTTCT TTCTTTCTTT CTTTCTTCTC TCTCTCCCTG TTCTTTTTCA

13801 CTTTTCTCCC ACCCAGCTCC CTCCATCTTG ATTCTCACCA ACCCATTTAT

13851 CTTTTCCTTT ACCTTTTTCT TTACCTTTTT CTTTATTTGG GGGCGGTGGG

13901 TGGGGAGGAC TATTTTTCTT TCCCTTACCC AAGTCTTAGA GCTGTCATTG

13951 GCTCCCACTT TACTCTTTCA TGCTACTGGT CTCCTCGGGG CCAGAGGAGT

14001 TACAGACCCT TGTTACTAGA ATTTGCAAGA AAAAAAAGAA CGTTCTCCCA

14051 AAGATGATGC TCTGCTCCCA AGGTCTACGG TAGTGACAAA GTTTGGAGAC

14101 TTTTTTTAAA GTAAGCATAA GTTAGATGCA TCTGAGGAAT AGATCGACGT

14151 TCAGGATATT AGGAATGACT ATAGAAGGCA AAGGTAAGAT AAAAGGAAAA

14201 TCAATAGGAT TAAAAAGCTA CTATACACAA TAACATATGC TGATTACATA

14251 CAGAAAATAA TGTAAGCATT TGAGTAGGCC ACTCAGCAAT GAGAACCTGA

14301 AGTGATATAA TAGGATTTAT TTTTTTGAAA GATTAAGGCA AAGGAGGCGG

14351 GGCACGGTGG CTCACGCCTA TAATCCCAGC CCTTTGGGAG GCCGAAGAGG

14401 GGGATCACTG GAGGTCAGGA GTTCGAGACC ACTCTGGCCA ACATGGTGAA

14451 AACGCATCTC TACTAAAACA AAACAAAACA AAACAAAATT AGCCGGGCAT

14501 GGTGGCGATG TGCTTGTAAT CCCAGCTACT GAAGAGGATG AGGTGGGAGG

14551 ATCACTTGAA CCCGGGAGTG GGAGGTTTCA GTAAGCCGAG TTTGCGCCAC

14601 TGCACTCCAG CCTGGGTGAC AGAGCGGGAC TCAGTGTCAA AAAATAAATA

14651 AATAAACAAA AAATAAAGGC AAAGGCTCAT AATGTCTCAT TTTGGCCCTA

14701 ATTTAGTCTT CAGTTAATTT CACCAAATAT TTTAAATAAC GCAAATTCAT

14751 TTCATGATAG TAAGAACACT GTACCTTCAA GTAAATCTTG TTTTTTATAT

14801 TCACGCCACA GATTTTTTTT TAATGAGACA AATTTGTGGT CCTGAATGAC

14851 AAAAACCTTC TTCTGTAATT CCCTCCACAC TGTATCCATG GCATGTGTAT

14901 TTTAAACATG CAGAGCAGGT ATATAAATAG CCCTGATTTT GCTAACAAAG

14951 CTAGTCAGCA TACCAACCTC ATTTTATACA ATTCCTTTCA ATGCATCTCA

15001 AGGTCGTTAC AGACTCAAGG ACAAAGGTGC TAAGACTTTA GAATTGTAAA

15051 TGATTCCGGG AGAAATTAGT AAGAACTAGG TATACAACAC AATTACCTCA

15101 ATACTAGAAA CTGAAGAGAA AGTGGCATCC TTCACAATAT AAAAAAAGCT

15151 GCTTAAAGGA AAAACACTCT GTGAGATAAA CTCTAAGCAT TTACAATGAA

Figure 5 (continued) (SEQ ID NO:10)

15201 AGGAATGCAA TAGTAGGAAG GCCAGTGAAG TCATAAACTG GAGACTTGCC

15251 TTTAAAACAT AAAAGGAAAT CTTTTGTACT ATAACTAGCT ACCATCAACA

15301 GGCAAATCTT GTAAAACCAA AATATTAGAT TACATTGTTA TTGAATTGAT

15351 CTCACCAGTT TGGATTATAG GAGTTTTGTC AGCTTCGGGT CAAATCCTGC

15401 TTGCCACCCA AGATGCAAAC CTAAAATAAG GCAGGCTAAT AAAATGTATT

15451 TATAAAAAAG AAAACAAATG TAAATCACTT AGGGTTCCAG AAGCTCAAGT

15501 TAAAACATAC ACAGACCCAG GGCACCCTGA AAATAACGAA GCCCCTTTCT

15551 GACTTTGAGT GAAATGGTAA CAGAGAAGGC TCCTACTTTA GTAAGCAGTC

15601 TTGCTGTTAG ACCCAGGAAG GAGGAAGGAT CAGTTGTTAA CTCCCCCAAG

15651 GGCACCCACT GAATGCAAGG CACAGAGCTA GTTCACCAGC AGCCCTAGGA

15701 TTAAAATTGT TACACTCAAA ATATATTGCT TTCCAGCATC CCATAATGTC

15751 TCACTTGTTC TACTAATTTC TCTGCCAATC CCTACCCTAA GCTACTTTCT

15801 TCTTTC**AG**AA TGAAGAAATT ACCACAACTT TCTTTGAACA GTCAGTGATT

15851 TGGGTCCCAG CAGAAAAGCC TATTGAAAAC CGAGATTTTC TTAAAAATTC

15901 CAAAATTCTG GAGATTTGTG ATAACGTGAC CATGTATTGG ATCAATCCCA

15951 CTCTAATATC AG**GT**ATGACA TTCTTATCCT CATCCTCCTC CTATTTTCTA

16001 AGACAGTAAC AGGTTAATGT TCCCAGCATT TAGAGGCTAC ATTTCTAGGA

16051 AAACAAATGA TCGGTTTATA TCTTCTTGGA TGAGTCTATA TATGCTGACT

16101 GCCAGTGAGG AGGGAATTTT TCACAGCCAA GTTATAAATC AGATAAATTA

16151 ATGTTGACAG GCTTTTAAAT ATCCCAGCAA AGGTAACACC ACCCTAATGT

16201 GACTAGTGAT AGCAAAAAAC TAATCTGTCT CTCTTTGCAC ATCTATATTC

16251 ATTGCAGTCG GTATGCCAGA CTATTGCTCA AACCACTGTG GGTCCAACCC

16301 AAGAATAGAG GAGAGAAGGG CAAAAACAAT GCTCTGCTAT CAAGAAGGAT

16351 TCCCTATGTC AAATGGTGGC ATTTTCCTTA GAAGATAGGA CAATCACTGC

16401 AATACAAATT TCTTTGTTGC TGTTTTAGGA ATTACTATAA AAATAGTTGA

16451 GGGCTTGATG TCCAGAGTAG AAGAGTCCAG CCCTCAACTT CCAGTCAGCC

16501 CATAGAGAAG GTAAAGGCCC CTGGATGTCT TAGTCTGTTT AGCTAACATG

16551 GAATTTGGCC TCCCTCATTC TGCACAGCAT AGATTACTCT ATTACCCAAG

16601 AGAGGTTATT CTCACCAATG TAAGCATCTG TGGGTACTCT TTAGAGATAT

16651 TAAAAAGGTG ACTAAATGCA GTGTGGTATC CTGGACTGGA TGCTGGAACA

16701 GAAAAAAACA TTAGTAGAAA TACTAATGAA ATCTGAATAA AGTCTGTAGT

Figure 5 (continued) (SEQ ID NO: 10)

16751 TTAGTTAATA ACATTGTACC AATTCTAACT TCTGATTTTT GATCATTGTA

16801 CCATGGTTAT GTGAGGTAAC ATTAGGGGAA ACCGGGTGAA GAATATGTGA

16851 ACTCTGTACT ATTTCTGCAA CTCTTCCATA AATCTAAGAT TATTTCTAAA

16901 CAGAAAGTTT AAAAAAAATT TAAGACAAGA AACATAAGGG AAGTGGGAGG

16951 ACAGAGGATG TCAAATCAAT TCAGGCAACC AAGAAATATT TATTGAGCAT

17001 CTACATGGGT AAAATCATCA TTGTAGGAGA TACAGAAGTC CCTTGTCCTC

17051 AGTTTTTAAT TTAGTGTGCT ATTAAGTCAA AGCAGTAAAT GTTTGATGCC

17101 AAAGAAGCAA AGAAGTTTGT GTATGTAGAA ATGATCAGGC TTGTTTTGGA

17151 AACTACTCTA AAATATCAGC TGAAGGAGCT TGGAAAAGTG GTCCATGCCT

17201 AAATGATGCC AGTTAGAAGA CACAATGAAA ATCTCTATCC CCAGCTGCTA

17251 TTCTCCCAAG CACTTCTTTC AGGAACATTT GCAAAGCAGT TGCATCACAA

17301 CTTTGCATTT ATTATCTT**AG** TTTCTGAGTT ACAAGACTTT GAGGAGGAGG

17351 GAGAAGATCT TCACTTTCCT GCCAACGAAA AAAAAGGGAT TGAACAAAAT

17401 GAACAGTGGG TGGTCCCTCA AGTGAAAGTA GAGAAGACCC GTCACGCCAG

17451 ACAAGCAAGT GAGGAAGAAC TTCCAATAAA TGACTAT**GT**G AGTTATGTTT

17501 ATGTAAACTC TTGATAGAGG AGGAAAAAAG AGCCCTCACA AAACTTACTA

17551 CCCCTCAGAT CACAATCATT GGAAGGAGAA CGATCATGGC TTTAACAAAA

17601 AAAAAAAAAT GTTTAAGAAA CTTGGAATCT CCAGGAGGCT CATTCAAAAT

17651 CAATCTCGAT CTCTGTAGAT ATATCTCCAT GATCAGGCTA AGGAGCTTAA

17701 TTTCTCACAA GCCTCCAACT CATACAGTCT TCTAGTTCTC CCCTGCTTGG

17751 TGAGAACTCT TGCTAGTGAA CCTTCCTCTC CCACCCCCAA CACTGGCTTC

17801 TTCTTCTTTC **AG**ACTGAAAA TGGAATAGAA TTTGATCCCA TGCTGGATGA

17851 GAGAGGTTAT TGTTGTATTT ACTGCCGTCG AGGCAACCGC TATTGCGGCC

17901 GCGTCTGTGA ACCTTTACTA GGCTACTACC CATATCCATA CTGCTACCAA

17951 GGAGGACGAG TCATCTGTCG TGTCATCATG CCTTGTAACT GGTGGGTGGC

18001 CCGCATGCTG GGGAGGGTCT AATAGGAGGT TTGAGCTCAA ATGCTTAAAC

18051 TGCTGGCAAC ATATAATAAA TGCATGCTAT TCAATGAATT TCTGCCTATG

18101 AGGCATCTGG CCCCTGGTAG CCAGCTCTCC AGAATTACTT GTAGGTAATT

18151 CCTCTCTTCA TGTTCTAATA AACTTCTACA TTATCACCAA CAGCCTGATT

18201 GCTGCTGAAC ACACTGGGTA AGTGTTTGCT GAGAATATTC ACAGCAACTG

18251 AAATGGAGCA GACCCGTCAA ACATAGATGA CTATGCTGAG GCTCAGAAAA

Figure 5 (continued) (SEQ ID NO: 10)

18301 CAATGCCCCA AAATGAAAGC CTCAGAAGCA AAAGTTTTTC TCTGACCTTC

18351 TCCTACCTGC CAGTCTCTCA GTCCCATTCT CCCCCGAGGC TAGCCATAAA

18401 AGCTAGAATC CCTCTTCCTC AAGGCGTGTC ATATAAATCA AAACCCCTTT

18451 TTTCCAAAGC CAGCCTTAAA CCTAAAAATA TCACTCTGAT TTTTCCTCTG

18501 CTTTTCTGTG TAAAAGCTGG CCATAAGGAA ATTATCTGAC CTACCTTGTT

18551 TGACTGTAGG TCATAAGACT CCCTTTCCAG AGAGGATCCA ACACAACAGG

18601 TTTGGAAGTA GAAAAGGGCA AAGCATAATG ATACCTAAAT ATCATATTTT

18651 ACTCTGCTGT ATTAAGACAG TGGAAATAAG GAGCTGATGA TAAAATGTCT

18701 CCATAACTGA TGCATTAAAA ACATATCCTT TTGTGCATAG CCCCAGTGTT

18751 TAATTATATG AAAGATATAT TGTGAACTTC CTTAAGTATC AATTAAGCTG

18801 CTGGACAGGC AGTGAGTCAT AGCAGGCCAA CACTAAAATA TTCCAAATGT

18851 GCAGACTCAG CAAAGGAGAA AGCTAAACTA GCAAAAAGGG TCTCAGGCTC

18901 TCCTCCTATA AACAGGAACT CAACTTAGGA CAAATTCAAC ATTCCATAGC

18951 CTTCACATCT GGCTCCTCTT TATATCAGTG TTTGATTAGA TACCTTTCCA

19001 AGGCAATTGT CTAAAAAACA AGAACCCATA AATACCATTC ACAGAAACAG

19051 GCATCATAGG AAATTATTGA GATGAGGAAT CAAATGGGAT TTTCCCTCCC

19101 AAATCACAGG GAGCTAGCAG TATACCAGTT TCTGAAAGAA AAGGTAAGGG

19151 CATACTTCCC AATCACTCCA TACTTTGCTT GGGAAAGGAA AAGAGGGCAA

19201 AGGGGAAAGA GAAAGAAGTG GTTAAAGGGT AACGTTAACC TTGATGTCCC

19251 TAGAGCCAGT TTCAGTTTCA AAATTCATGC TCTACCAATC TGTCACCCCT

19301 TCTAGTTTCA TTTGAAGGTA GCATATACTT CATTGCTCAA CTTCAAAGTT

19351 ATTGTGTCAT GAAAATGTCT GTCTGGAACT TTCTATCAGA AAAATATTTT

19401 CAAATAAATT CTTCACCAAG ATGCCTGTGG GCCAGCAGTC ATTTTCTCCC

19451 ATTTAAGAGT TGAGGAAATC CATAGAAAAT GGCTTTGCCC AAGGTTGAGT

19501 TGTGATTAGC CAGTCATTAA TTGGTAATTA GCATATATGC CATACTTAGA

19551 TTACCAATTT TCACAGAAAG AGAGCACAAT AAAATATATC TAACAATGCA

19601 CATTACTGAA AATATTTTTA ATTGATGATT TAGAAGACAT CTGATTGTTT

19651 TAGAGTACTT CGATTATCAT AACAAGCATT AATTTATAGT TGAAATAGAA

19701 AAACGTCAAA GATAGGAAAT TACTTTGAGG TTATATGAAC TTTATGATGT

19751 TTCTATCATT GTTTTTTAAT AAGACTACTC ATTTTTAACA TTGCTTTTTT

19801 AGATCAGCCA ATAATTTTAT TATATTTAAT TTTTTCTATT TTTTATTTAT

Figure 5 (continued) (SEQ ID NO: 10)

19851 TTTTCCATAG GTTATTGGGG TACAGGAGGT ATTTGGTTAC ATGAGTAAGT

19901 TATTTAGTGA TGATTTGTGA GATTTTGGTG CACCCATCAC CCAAGCAGTA

19951 TACACTGAAC CCTATTTGTG GTCTTTTATC CCTCGGCCCC CCACCCCACC

20001 ATTTCTCCCA AGTCACCAAA GTCCATTCTG TATTTTTTTT TTTTTTTGAG

20051 ACGGAGTCTT CCTTTGTCGC CCAGGCTGGA GTGCAGTGGT GCGATCTCTG

20101 CTCACTGCAA GCTCCGCCTC CCGAGTTCAC GCCATTCTCC TGCCTCAGCC

20151 TCCCGAGTAG CTGCGACTAC AGGTGCCTGC CACCACGCCC GGCTAATTTT

20201 TTTTTTTTTT TTTTGTATTT TTAGTAGAGA CAGGGTTTCA CCGTCTTAGT

20251 CAGGATCGTC TCGATCTCCT GACCTCAGGT GATCCACCCG CCTCGGCCTC

20301 TCAAAGTGCT GGGATTACAG GAGTGAGCCA CCGCGCCTGG CCCATTGTAT

20351 CATTCTTATG CCTTTGCGTC CTCATAGCTT AGCTTCCACG TATCTGTGAG

20401 AACATGCAAT GTTTGGTTTT CCATTCTTGA GTTACTTCAC TTAGAATAAT

20451 AGTCTCCAGC CGGGCACGGT GGCTAAGCCT GTAATGACAG CACTTTTGGG

20501 AGGCCAAGGT GGGCGGATCA CGAGGTCAGG AGATCAAGAC CATCCTGGCT

20551 AACACGGTGA AACCCCGTCT CTACTAAAAA TACAAAAAAA TTAGCCAGGT

20601 GTGGTGGCAG GCGCCTGTAG ACCCAGCCAC TCAGGAGGCT GAGGCAGGAG

20651 AATGGCATGA ACCCGGGAGG CGGAGCTTGC ACTGAGCCGA GATCGCACCA

20701 CTGCACTCCA GCCTGGGCGA CAGATCAAGA CTCTGTCTCA AAAAAAAAAA

20751 AGAAAGAATA ATAGTCTCCA ATCTCATCCA TGTTACTGCA AATGCCATTA

20801 ATTCGTTCCT TGTTATGACT GAGTAGTATT CCATCATACA TATGTGTGTG

SEQUENCE LISTING

<110> Amgen, Inc.

<120> Chondromodulin-I Related Peptide

<130> 01017/35944

<140>
<141>

<150> US 09/724,310
<151> 2000-11-28

<150> US 60/176,898
<151> 2000-01-19

<160> 20

<170> PatentIn Ver. 2.0

<210> 1
<211> 1206
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (71)..(1021)

<400> 1
gcaactccac ctcagcagtg gtctctcagt cctctcaaag caaggaaaga gtactgtgtg 60

ctgagagacc atg gca aag aat cct cca gag aat tgt gaa gac tgt cac 109
Met Ala Lys Asn Pro Pro Glu Asn Cys Glu Asp Cys His
1 5 10

att cta aat gca gaa gct ttt aaa tcc aag aaa ata tgt aaa tca ctt 157
Ile Leu Asn Ala Glu Ala Phe Lys Ser Lys Lys Ile Cys Lys Ser Leu
15 20 25

aag att tgt gga ctg gtg ttt ggt atc ctg gcc cta act cta att gtc 205
Lys Ile Cys Gly Leu Val Phe Gly Ile Leu Ala Leu Thr Leu Ile Val
30 35 40 45

ctg ttt tgg ggg agc aag cac ttc tgg ccg gag gta ccc aaa aaa gcc 253
Leu Phe Trp Gly Ser Lys His Phe Trp Pro Glu Val Pro Lys Lys Ala
50 55 60

tat gac atg gag cac act ttc tac agc aat gga gag aag aag aag att 301
Tyr Asp Met Glu His Thr Phe Tyr Ser Asn Gly Glu Lys Lys Lys Ile
65 70 75

tac atg gaa att gat cct gtg acc aga act gaa ata ttc aga agc gga 349
Tyr Met Glu Ile Asp Pro Val Thr Arg Thr Glu Ile Phe Arg Ser Gly
80 85 90

aat ggc act gat gaa aca ttg gaa gtg cac gac ttt aaa aac gga tac 397
Asn Gly Thr Asp Glu Thr Leu Glu Val His Asp Phe Lys Asn Gly Tyr
95 100 105

```
act ggc atc tac ttc gtg ggt ctt caa aaa tgt ttt atc aaa act cag   445
Thr Gly Ile Tyr Phe Val Gly Leu Gln Lys Cys Phe Ile Lys Thr Gln
110                 115                 120                 125

att aaa gtg att cct gaa ttt tct gaa cca gaa gag gaa ata gat gag   493
Ile Lys Val Ile Pro Glu Phe Ser Glu Pro Glu Glu Glu Ile Asp Glu
                130                 135                 140

aat gaa gaa att acc aca act ttc ttt gaa cag tca gtg att tgg gtc   541
Asn Glu Glu Ile Thr Thr Thr Phe Phe Glu Gln Ser Val Ile Trp Val
            145                 150                 155

cca gca gaa aag cct att gaa aac cga gat ttt ctt aaa aat tcc aaa   589
Pro Ala Glu Lys Pro Ile Glu Asn Arg Asp Phe Leu Lys Asn Ser Lys
        160                 165                 170

att ctg gag att tgt gat aac gtg acc atg tat tgg atc aat ccc act   637
Ile Leu Glu Ile Cys Asp Asn Val Thr Met Tyr Trp Ile Asn Pro Thr
    175                 180                 185

cta ata tca gtt tct gag tta caa gac ttt gag gag gag gga gaa gat   685
Leu Ile Ser Val Ser Glu Leu Gln Asp Phe Glu Glu Glu Gly Glu Asp
190                 195                 200                 205

ctt cac ttt cct gcc aac gaa aaa aaa ggg att gaa caa aat gaa cag   733
Leu His Phe Pro Ala Asn Glu Lys Lys Gly Ile Glu Gln Asn Glu Gln
                210                 215                 220

tgg gtg gtc cct caa gtg aaa gta gag aag acc cgt cac gcc aga caa   781
Trp Val Val Pro Gln Val Lys Val Glu Lys Thr Arg His Ala Arg Gln
            225                 230                 235

gca agt gag gaa gaa ctt cca ata aat gac tat act gaa aat gga ata   829
Ala Ser Glu Glu Glu Leu Pro Ile Asn Asp Tyr Thr Glu Asn Gly Ile
        240                 245                 250

gaa ttt gat ccc atg ctg gat gag aga ggt tat tgt tgt att tac tgc   877
Glu Phe Asp Pro Met Leu Asp Glu Arg Gly Tyr Cys Cys Ile Tyr Cys
    255                 260                 265

cgt cga ggc aac cgc tat tgc cgc cgc gtc tgt gaa cct tta cta ggc   925
Arg Arg Gly Asn Arg Tyr Cys Arg Arg Val Cys Glu Pro Leu Leu Gly
270                 275                 280                 285

tac tac cca tat cca tac tgc tac caa gga gga cga gtc atc tgt cgt   973
Tyr Tyr Pro Tyr Pro Tyr Cys Tyr Gln Gly Gly Arg Val Ile Cys Arg
                290                 295                 300

gtc atc atg cct tgt aac tgg tgg gtg gcc cgc atg ctg ggg agg gtc   1021
Val Ile Met Pro Cys Asn Trp Trp Val Ala Arg Met Leu Gly Arg Val
            305                 310                 315

taataggagg tttgagctca aatgcttaaa ctgctggcaa catataataa atgcatgcta 1081

ttcaatgaat ttctgcctat gaggcatctg gcccctggta gccagctctc cagaattact 1141

tgtaggtaat tcctctcttc atgttctaat aaacttctac attatcacca aaaaaaaaaa 1201

aaaaa                                                             1206
```

<210> 2
<211> 317
<212> PRT
<213> Homo sapiens

<400> 2

```
Met Ala Lys Asn Pro Pro Glu Asn Cys Glu Asp Cys His Ile Leu Asn
 1               5                  10                  15

Ala Glu Ala Phe Lys Ser Lys Lys Ile Cys Lys Ser Leu Lys Ile Cys
            20                  25                  30

Gly Leu Val Phe Gly Ile Leu Ala Leu Thr Leu Ile Val Leu Phe Trp
        35                  40                  45

Gly Ser Lys His Phe Trp Pro Glu Val Pro Lys Lys Ala Tyr Asp Met
    50                  55                  60

Glu His Thr Phe Tyr Ser Asn Gly Glu Lys Lys Lys Ile Tyr Met Glu
65                  70                  75                  80

Ile Asp Pro Val Thr Arg Thr Glu Ile Phe Arg Ser Gly Asn Gly Thr
                85                  90                  95

Asp Glu Thr Leu Glu Val His Asp Phe Lys Asn Gly Tyr Thr Gly Ile
            100                 105                 110

Tyr Phe Val Gly Leu Gln Lys Cys Phe Ile Lys Thr Gln Ile Lys Val
        115                 120                 125

Ile Pro Glu Phe Ser Glu Pro Glu Glu Glu Ile Asp Glu Asn Glu Glu
    130                 135                 140

Ile Thr Thr Thr Phe Phe Glu Gln Ser Val Ile Trp Val Pro Ala Glu
145                 150                 155                 160

Lys Pro Ile Glu Asn Arg Asp Phe Leu Lys Asn Ser Lys Ile Leu Glu
                165                 170                 175

Ile Cys Asp Asn Val Thr Met Tyr Trp Ile Asn Pro Thr Leu Ile Ser
            180                 185                 190

Val Ser Glu Leu Gln Asp Phe Glu Glu Glu Gly Glu Asp Leu His Phe
        195                 200                 205

Pro Ala Asn Glu Lys Lys Gly Ile Glu Gln Asn Glu Gln Trp Val Val
    210                 215                 220

Pro Gln Val Lys Val Glu Lys Thr Arg His Ala Arg Gln Ala Ser Glu
225                 230                 235                 240

Glu Glu Leu Pro Ile Asn Asp Tyr Thr Glu Asn Gly Ile Glu Phe Asp
                245                 250                 255

Pro Met Leu Asp Glu Arg Gly Tyr Cys Cys Ile Tyr Cys Arg Arg Gly
            260                 265                 270

Asn Arg Tyr Cys Arg Arg Val Cys Glu Pro Leu Leu Gly Tyr Tyr Pro
        275                 280                 285
```

```
Tyr Pro Tyr Cys Tyr Gln Gly Gly Arg Val Ile Cys Arg Val Ile Met
    290                 295                 300

Pro Cys Asn Trp Trp Val Ala Arg Met Leu Gly Arg Val
305                 310                 315


<210> 3
<211> 951
<212> DNA
<213> mouse

<220>
<221> CDS
<222> (1)..(951)

<400> 3
atg gca aag aat cct cca gag aac tgt gag ggc tgt cac att cta aat   48
Met Ala Lys Asn Pro Pro Glu Asn Cys Glu Gly Cys His Ile Leu Asn
  1               5                  10                  15

gca gaa gct ctg aaa tct aag aag ata tgt aaa tca ctg aag att tgt   96
Ala Glu Ala Leu Lys Ser Lys Lys Ile Cys Lys Ser Leu Lys Ile Cys
             20                  25                  30

gga cta gtg ttt ggt atc ctg gcc tta act cta att gtc ctg ttt tgg   144
Gly Leu Val Phe Gly Ile Leu Ala Leu Thr Leu Ile Val Leu Phe Trp
         35                  40                  45

ggg agc aaa cac ttc tgg ccc gag gta tcc aag aaa acc tat gac atg   192
Gly Ser Lys His Phe Trp Pro Glu Val Ser Lys Lys Thr Tyr Asp Met
     50                  55                  60

gag cac act ttc tac agc aac ggc gag aag aag aag att tac atg gaa   240
Glu His Thr Phe Tyr Ser Asn Gly Glu Lys Lys Lys Ile Tyr Met Glu
 65                  70                  75                  80

att gat ccc ata acc aga aca gaa ata ttc aga agt gga aat ggc act   288
Ile Asp Pro Ile Thr Arg Thr Glu Ile Phe Arg Ser Gly Asn Gly Thr
                 85                  90                  95

gat gaa aca ttg gaa gtc cat gac ttt aaa aat gga tac act ggc atc   336
Asp Glu Thr Leu Glu Val His Asp Phe Lys Asn Gly Tyr Thr Gly Ile
            100                 105                 110

tac ttt gta ggt ctt caa aaa tgc ttt att aaa act caa atc aaa gtg   384
Tyr Phe Val Gly Leu Gln Lys Cys Phe Ile Lys Thr Gln Ile Lys Val
        115                 120                 125

att cct gaa ttt tct gaa cca gag gaa gaa ata gat gag aat gaa gaa   432
Ile Pro Glu Phe Ser Glu Pro Glu Glu Glu Ile Asp Glu Asn Glu Glu
    130                 135                 140

att act aca act ttc ttt gaa cag tca gtg att tgg gtt ccc gca gaa   480
Ile Thr Thr Thr Phe Phe Glu Gln Ser Val Ile Trp Val Pro Ala Glu
145                 150                 155                 160

aag cct att gaa aac aga gac ttc ctg aaa aat tct aaa att ctg gag   528
Lys Pro Ile Glu Asn Arg Asp Phe Leu Lys Asn Ser Lys Ile Leu Glu
                165                 170                 175
```

```
att tgc gat aat gtg acc atg tac tgg atc aat ccc act cta ata gca   576
Ile Cys Asp Asn Val Thr Met Tyr Trp Ile Asn Pro Thr Leu Ile Ala
            180                 185                 190

gtt tca gaa tta cag gac ttt gag gag gac ggt gaa gat ctt cac ttt   624
Val Ser Glu Leu Gln Asp Phe Glu Glu Asp Gly Glu Asp Leu His Phe
        195                 200                 205

cct acc agt gaa aaa aag ggg att gac cag aat gag caa tgg gtg gtc   672
Pro Thr Ser Glu Lys Lys Gly Ile Asp Gln Asn Glu Gln Trp Val Val
    210                 215                 220

ccg caa gtg aag gtg gag aag acc cgc cac acc aga caa gca agc gag   720
Pro Gln Val Lys Val Glu Lys Thr Arg His Thr Arg Gln Ala Ser Glu
225                 230                 235                 240

gaa gac ctt cct ata aat gac tat act gaa aat gga att gaa ttt gac   768
Glu Asp Leu Pro Ile Asn Asp Tyr Thr Glu Asn Gly Ile Glu Phe Asp
                245                 250                 255

cca atg ctg gat gag aga ggt tac tgt tgt att tac tgt cgt cga ggc   816
Pro Met Leu Asp Glu Arg Gly Tyr Cys Cys Ile Tyr Cys Arg Arg Gly
            260                 265                 270

aac cgt tac tgc cgc cgt gtc tgt gaa cct tta cta ggc tac tac cca   864
Asn Arg Tyr Cys Arg Arg Val Cys Glu Pro Leu Leu Gly Tyr Tyr Pro
        275                 280                 285

tac ccc tac tgc tac caa gga ggt cga gtc atc tgt cgt gtc atc atg   912
Tyr Pro Tyr Cys Tyr Gln Gly Gly Arg Val Ile Cys Arg Val Ile Met
    290                 295                 300

cct tgc aac tgg tgg gtg gcc cgc atg ctt ggg aga gtc               951
Pro Cys Asn Trp Trp Val Ala Arg Met Leu Gly Arg Val
305                 310                 315
```

<210> 4
<211> 317
<212> PRT
<213> mouse

<400> 4

```
Met Ala Lys Asn Pro Pro Glu Asn Cys Glu Gly Cys His Ile Leu Asn
  1               5                  10                  15

Ala Glu Ala Leu Lys Ser Lys Lys Ile Cys Lys Ser Leu Lys Ile Cys
             20                  25                  30

Gly Leu Val Phe Gly Ile Leu Ala Leu Thr Leu Ile Val Leu Phe Trp
         35                  40                  45

Gly Ser Lys His Phe Trp Pro Glu Val Ser Lys Lys Thr Tyr Asp Met
     50                  55                  60

Glu His Thr Phe Tyr Ser Asn Gly Glu Lys Lys Lys Ile Tyr Met Glu
 65                  70                  75                  80

Ile Asp Pro Ile Thr Arg Thr Glu Ile Phe Arg Ser Gly Asn Gly Thr
                 85                  90                  95

Asp Glu Thr Leu Glu Val His Asp Phe Lys Asn Gly Tyr Thr Gly Ile
             100                 105                 110
```

```
Tyr Phe Val Gly Leu Gln Lys Cys Phe Ile Lys Thr Gln Ile Lys Val
        115                 120                 125

Ile Pro Glu Phe Ser Glu Pro Glu Glu Glu Ile Asp Glu Asn Glu Glu
    130                 135                 140

Ile Thr Thr Thr Phe Phe Glu Gln Ser Val Ile Trp Val Pro Ala Glu
145                 150                 155                 160

Lys Pro Ile Glu Asn Arg Asp Phe Leu Lys Asn Ser Lys Ile Leu Glu
                165                 170                 175

Ile Cys Asp Asn Val Thr Met Tyr Trp Ile Asn Pro Thr Leu Ile Ala
            180                 185                 190

Val Ser Glu Leu Gln Asp Phe Glu Glu Asp Gly Glu Asp Leu His Phe
        195                 200                 205

Pro Thr Ser Glu Lys Lys Gly Ile Asp Gln Asn Glu Gln Trp Val Val
    210                 215                 220

Pro Gln Val Lys Val Glu Lys Thr Arg His Thr Arg Gln Ala Ser Glu
225                 230                 235                 240

Glu Asp Leu Pro Ile Asn Asp Tyr Thr Glu Asn Gly Ile Glu Phe Asp
                245                 250                 255

Pro Met Leu Asp Glu Arg Gly Tyr Cys Cys Ile Tyr Cys Arg Arg Gly
            260                 265                 270

Asn Arg Tyr Cys Arg Arg Val Cys Glu Pro Leu Leu Gly Tyr Tyr Pro
        275                 280                 285

Tyr Pro Tyr Cys Tyr Gln Gly Gly Arg Val Ile Cys Arg Val Ile Met
    290                 295                 300

Pro Cys Asn Trp Trp Val Ala Arg Met Leu Gly Arg Val
305                 310                 315
```

<210> 5
<211> 334
<212> PRT
<213> Mouse

<400> 5

```
Met Thr Glu Asn Ser Asp Lys Val Pro Ile Thr Met Val Gly Pro Glu
  1               5                  10                  15

Asp Val Glu Phe Cys Ser Pro Pro Ala Tyr Thr Thr Val Thr Val Lys
             20                  25                  30

Pro Ser Gly Ser Pro Thr Arg Leu Leu Lys Val Gly Ala Val Val Leu
         35                  40                  45

Ile Ser Gly Ala Val Leu Leu Leu Phe Gly Ala Ile Gly Ala Phe Tyr
     50                  55                  60

Phe Trp Lys Gly Asn Asp Asn His Ile Tyr Asn Val His Tyr Ser Met
 65                  70                  75                  80

Ser Ile Asn Gly Lys Leu Gln Asp Gly Ser Met Glu Ile Asp Ala Val
                 85                  90                  95
```

```
Asn Asn Leu Glu Thr Phe Lys Met Gly Ser Gly Ala Lys Glu Ala Ile
            100                 105                 110

Glu Val Asn Asp Phe Lys Asn Gly Ile Thr Gly Ile Arg Phe Ala Gly
        115                 120                 125

Gly Glu Lys Cys Tyr Ile Lys Ala Gln Val Lys Ala Arg Ile Pro Glu
    130                 135                 140

Val Gly Thr Val Thr Lys Gln Ser Ile Ser Glu Leu Glu Gly Lys Ile
145                 150                 155                 160

Met Pro Val Asn Tyr Glu Glu Asn Ser Leu Ile Trp Val Ala Val Asp
                165                 170                 175

Gln Pro Val Lys Asp Ser Ser Phe Leu Ser Ser Lys Ile Leu Glu Leu
            180                 185                 190

Cys Gly Asp Leu Pro Ile Phe Trp Leu Lys Pro Met Tyr Pro Lys Glu
        195                 200                 205

Ile Gln Arg Glu Arg Arg Glu Val Val Arg Asn Ser Ala Pro Ser Thr
    210                 215                 220

Thr Arg Arg Pro His Ser Glu Pro Arg Gly Asn Ala Gly Pro Gly Arg
225                 230                 235                 240

Leu Ser Asn Gly Thr Arg Pro Asn Val Gln Asp Asp Ala Glu Pro Phe
                245                 250                 255

Asn Pro Asp Asn Pro Tyr His Gln Gln Glu Gly Glu Ser Met Thr Phe
            260                 265                 270

Asp Pro Arg Leu Asp His Glu Gly Ile Cys Cys Ile Glu Cys Arg Arg
        275                 280                 285

Ser Tyr Thr His Cys Gln Lys Ile Cys Glu Pro Leu Gly Gly Tyr Tyr
    290                 295                 300

Pro Trp Pro Tyr Asn Tyr Gln Gly Cys Arg Ser Ala Cys Arg Val Val
305                 310                 315                 320

Met Pro Cys Ser Trp Trp Val Ala Arg Ile Leu Gly Met Val
                325                 330
```

<210> 6
<211> 334
<212> PRT
<213> Rat

<400> 6

```
Met Thr Glu Asn Ser Asp Lys Val Pro Ile Thr Met Val Gly Pro Glu
  1               5                  10                  15

Asp Val Glu Phe Cys Ser Pro Pro Ala Tyr Ala Thr Val Thr Val Lys
             20                  25                  30

Pro Ser Gly Ser Pro Thr Arg Leu Leu Lys Val Gly Ala Val Val Leu
         35                  40                  45
```

```
Ile Ser Gly Ala Val Leu Leu Leu Phe Gly Ala Ile Gly Ala Phe Tyr
     50                  55                  60

Phe Trp Lys Gly Asn Asp Asn His Ile Tyr Asn Val His Tyr Thr Met
 65                  70                  75                  80

Ser Ile Asn Gly Arg Leu Gln Asp Ala Ser Met Glu Ile Asp Ala Ala
                 85                  90                  95

Asn Asn Leu Glu Thr Phe Lys Met Gly Ser Gly Ala Glu Glu Ala Ile
            100                 105                 110

Glu Val Asn Asp Phe Gln Asn Gly Ile Thr Gly Ile Arg Phe Ala Gly
        115                 120                 125

Gly Glu Lys Cys Tyr Ile Lys Ala Gln Val Lys Ala Arg Ile Pro Glu
    130                 135                 140

Val Ser Thr Gly Thr Lys Gln Ser Ile Ser Glu Leu Glu Gly Lys Ile
145                 150                 155                 160

Met Pro Val Lys Tyr Glu Glu Asn Ser Leu Ile Trp Val Ala Val Asp
                165                 170                 175

Gln Pro Val Lys Asp Asn Ser Phe Leu Ser Ser Lys Ile Leu Glu Phe
            180                 185                 190

Cys Gly Asp Leu Pro Ile Phe Trp Leu Lys Pro Met Tyr Pro Lys Glu
        195                 200                 205

Ile Pro Arg Glu Arg Arg Glu Val Val Arg Ser Ser Ala Pro Ser Thr
    210                 215                 220

Thr Arg Arg Pro His Ser Glu Pro Arg Gly Asn Ala Gly Pro Gly Arg
225                 230                 235                 240

Leu Ser Asn Arg Thr Arg Pro Ser Val Gln Asp Asp Glu Glu Pro Phe
                245                 250                 255

Asn Pro Asp Asn Pro Tyr His Gln Gln Glu Gly Glu Ser Met Thr Phe
            260                 265                 270

Asp Pro Arg Leu Asp His Glu Gly Ile Cys Cys Ile Glu Cys Arg Arg
        275                 280                 285

Ser Tyr Thr His Cys Gln Lys Ile Cys Glu Pro Leu Gly Gly Tyr Tyr
    290                 295                 300

Pro Trp Pro Tyr Asn Tyr Gln Gly Cys Arg Ser Ala Cys Arg Val Val
305                 310                 315                 320

Met Pro Cys Ser Trp Trp Val Ala Arg Ile Leu Gly Met Val
                325                 330
```

<210> 7
<211> 335
<212> PRT
<213> Bovine

<400> 7

```
Met Thr Glu Asn Ser Asp Lys Val Pro Ile Ala Leu Val Gly Pro Asp
  1               5                  10                  15

Asp Val Glu Phe Cys Ser Pro Pro Ala Tyr Ala Ala Val Thr Val Lys
             20                  25                  30

Pro Ser Ser Pro Ala Arg Leu Leu Lys Val Gly Ala Val Val Leu Ile
         35                  40                  45

Ser Gly Ala Val Leu Leu Leu Leu Gly Ala Ile Gly Ala Phe Tyr Phe
     50                  55                  60

Trp Lys Gly Ser Asp Asn His Ile Tyr Asn Val His Tyr Thr Met Ser
 65                  70                  75                  80

Ile Asn Gly Lys Leu Gln Asp Gly Ser Met Glu Ile Asp Ala Gly Asn
                 85                  90                  95

Asn Leu Glu Thr Phe Lys Met Gly Ser Gly Ala Glu Glu Ala Val Glu
            100                 105                 110

Val Asn Asp Phe Gln Asn Gly Ile Thr Gly Ile Arg Phe Ala Gly Gly
        115                 120                 125

Glu Lys Cys Tyr Ile Lys Ala Gln Val Lys Ala Arg Ile Pro Glu Val
    130                 135                 140

Gly Thr Met Thr Lys Gln Ser Ile Ser Ser Glu Leu Glu Gly Lys Ile
145                 150                 155                 160

Met Pro Val Lys Tyr Glu Glu Asn Ser Leu Ile Trp Val Ala Gly Asp
                165                 170                 175

Gln Pro Val Lys Asp Asn Ser Phe Leu Ser Ser Lys Val Leu Glu Leu
            180                 185                 190

Cys Gly Asp Leu Pro Ile Phe Trp Leu Lys Pro Thr Tyr Pro Lys Glu
        195                 200                 205

Ile Gln Arg Glu Arg Arg Glu Leu Val Arg Lys Ile Val Thr Thr Thr
    210                 215                 220

Thr Thr Arg Arg Leu Arg Ser Gly Pro Gln Gly Thr Pro Ala Pro Gly
225                 230                 235                 240

Arg Pro Asn Asn Gly Thr Arg Pro Ser Val Gln Glu Asp Ala Glu Pro
                245                 250                 255

Phe Asn Pro Asp Asn Pro Tyr His Gln Gln Glu Gly Glu Ser Met Thr
            260                 265                 270

Phe Asp Pro Arg Leu Asp His Glu Gly Ile Cys Cys Ile Glu Cys Arg
        275                 280                 285

Arg Ser Tyr Thr His Cys Gln Lys Ile Cys Glu Pro Leu Gly Gly Tyr
    290                 295                 300

His Pro Trp Pro Tyr Asn Tyr Gln Gly Cys Arg Ser Ala Cys Arg Val
305                 310                 315                 320

Ile Met Pro Cys Ser Trp Trp Val Ala Arg Ile Leu Gly Met Val
                325                 330                 335
```

<210> 8
<211> 334
<212> PRT
<213> Homo sapiens

<400> 8

```
Met Thr Glu Asn Ser Asp Lys Val Pro Ile Ala Leu Val Gly Pro Asp
  1               5                  10                  15

Asp Val Glu Phe Cys Ser Pro Pro Ala Tyr Ala Thr Leu Thr Val Lys
             20                  25                  30

Pro Ser Ser Pro Ala Arg Leu Leu Lys Val Gly Ala Val Val Leu Ile
         35                  40                  45

Ser Gly Ala Val Leu Leu Leu Phe Gly Ala Ile Gly Ala Phe Tyr Phe
     50                  55                  60

Trp Lys Gly Ser Asp Ser His Ile Tyr Asn Val His Tyr Thr Met Ser
 65                  70                  75                  80

Ile Asn Gly Lys Leu Gln Asp Gly Ser Met Glu Ile Asp Ala Gly Asn
                 85                  90                  95

Asn Leu Glu Thr Phe Lys Met Gly Ser Gly Ala Glu Glu Ala Ile Ala
            100                 105                 110

Val Asn Asp Phe Gln Asn Gly Ile Thr Gly Ile Arg Phe Ala Gly Gly
        115                 120                 125

Glu Lys Cys Tyr Ile Lys Ala Gln Val Lys Ala Arg Ile Pro Glu Val
    130                 135                 140

Gly Ala Val Thr Lys Gln Ser Ile Ser Ser Lys Leu Glu Gly Lys Ile
145                 150                 155                 160

Met Pro Val Lys Tyr Glu Glu Asn Ser Leu Ile Trp Val Ala Val Asp
                165                 170                 175

Gln Pro Val Lys Asp Asn Ser Phe Leu Ser Ser Lys Val Leu Glu Leu
            180                 185                 190

Cys Gly Asp Leu Pro Ile Phe Trp Leu Lys Pro Thr Tyr Pro Lys Glu
        195                 200                 205

Ile Gln Arg Glu Arg Arg Glu Val Val Arg Lys Ile Val Pro Thr Thr
    210                 215                 220

Thr Lys Arg Pro His Ser Gly Pro Arg Ser Asn Pro Gly Ala Gly Arg
225                 230                 235                 240

Leu Asn Asn Glu Thr Arg Pro Ser Val Gln Glu Asp Ser Gln Ala Phe
                245                 250                 255

Asn Pro Asp Asn Pro Tyr His Gln Gln Glu Gly Glu Ser Met Thr Phe
            260                 265                 270

Asp Pro Arg Leu Asp His Glu Gly Ile Cys Cys Ile Glu Cys Arg Arg
        275                 280                 285
```

```
Ser Tyr Thr His Cys Gln Lys Ile Cys Glu Pro Leu Gly Gly Tyr Tyr
    290                 295                 300

Pro Trp Pro Tyr Asn Tyr Gln Gly Cys Arg Ser Ala Cys Arg Val Ile
305                 310                 315                 320

Met Pro Cys Ser Trp Trp Val Ala Arg Ile Leu Gly Met Val
                325                 330
```

<210> 9
<211> 234
<212> PRT
<213> Rabbit

<400> 9

```
Thr Phe Lys Met Gly Ser Gly Ala Glu Glu Ala Ile Glu Val Asn Asp
  1               5                  10                  15

Phe Gln Asn Gly Ile Thr Gly Ile Arg Phe Ala Gly Gly Glu Lys Cys
             20                  25                  30

Tyr Ile Lys Ala Gln Val Lys Ala Arg Val Pro Glu Val Gly Thr Val
         35                  40                  45

Thr Gln Gln Ser Ile Ser Ser Glu Leu Glu Gly Lys Ile Met Pro Val
     50                  55                  60

Lys His Glu Glu Glu Ala Leu Val Trp Val Ala Val Gly Gln Pro Val
 65                  70                  75                  80

Gln Asp Asn Ser Phe Leu Ser Ala Arg Val Leu Glu Leu Cys Gly Asp
                85                  90                  95

Leu Pro Ile Phe Trp Leu Lys Pro Thr Tyr Pro Lys Glu Ile Gln Arg
            100                 105                 110

Glu Arg Arg Glu Val Val Arg Lys Thr Val Pro Thr Thr Thr Lys Arg
        115                 120                 125

Pro His Ser Gly Pro Arg Gly Asn Pro Gly Pro Ala Arg Met Arg Asn
    130                 135                 140

Asp Ser Arg Pro Ser Val Gln Glu Asp Ser Glu Pro Phe Asn Pro Asp
145                 150                 155                 160

Asn Pro Tyr His Gln Glu Gly Glu Ser Met Thr Phe Asp Pro Arg Leu
                165                 170                 175

Asp His Glu Gly Ile Cys Cys Ile Glu Cys Arg Arg Ser Tyr Thr His
            180                 185                 190

Cys Gln Lys Ile Cys Glu Pro Leu Gly Gly Tyr Asn Pro Trp Pro Tyr
        195                 200                 205

Asn Tyr Gln Gly Cys Arg Ser Ala Cys Arg Val Val Met Pro Cys Ser
    210                 215                 220

Trp Trp Val Ala Arg Ile Leu Gly Met Val
225                 230
```

<210> 10
<211> 20850
<212> DNA
<213> Homo sapiens

<400> 10
gtgctgttcc tttggcctgg aataatcttt tccccactca ttaccagcta acttctactc 60

attattcagg tcttagctta agtgacaact cctcagcatg gatttccctc acccatccca 120

cccaatccaa attaggtgct ttaaagcacc tgcacttttc cttcttagca tatatcatca 180

tttatgtact tttatttaca ggatttttt ccttttagtt taatgtctgt ctacccaaca 240

agactctgag ctctataaag gaggaactgt gtttgcttgt ctactactga atatatatct 300

agctcctata acaatgcctg ccacataggg gacactgaat aaatacatgc taatcaatga 360

acagtctatt gaaatgccat gtcctttgtg ccatttttcc tgcttttctg aatcagaatt 420

aatcactctt ccctctatgt ttccatggaa tattgctttg actcttggtt taattcaatc 480

cttttaacaa gaaaggcatg ttttaaaaga tctcaaagtc tattcaggag aggaacacat 540

aaaaaagaat tataatacag catacaaagt ataacggttg aagtatgagc aagatacaaa 600

agtagagcaa gggaaagggt aattaactta acgaagtgga agagatgaga ccaggaaagg 660

ctttcagggg gtgataacat ctaggcagca ttttgaaagc caaatgggaa ttttccaggt 720

aaacaaggga aggcagaaaa ctccaggcaa tgaatatagc atgggcaaat gtcctacagt 780

gtgaagaagt agggcatgtt cattaatagg gtgtacagtg tgaggcaggc atcagaggca 840

tagtgtgaat aacatcactg aggggaacca tctgacccaa taaacataag tttctccatg 900

agactgctct tgtggctttt ctttttttctc tctttaagct tgaccaaaga attgtagaac 960

atatagaatc gtgacatatt tcataattta tgtaatcagt ttaatgcaca atcatgtatc 1020

caacaattag aaaattattt tccttgagac attaggatgt ggaggtagtc atatgtacat 1080

gccccatttt cctgcccagc tgcgacctcc cagttggctc catcgctgta gaaaaaaatt 1140

gtccaaaatc aagactagag agttggacaa agtcgagatc attaaatgcc tactatgtca 1200

caataagtgg cttggatttt atagacaatg aatagatact aaaggatttt aagcaaggga 1260

gagattatcc tcaaatttga gtcttagaaa gatcaccttg aagacaatgt atgagagaca 1320

aggaaaacag aattctattg caatagtaaa agtgactgag aacctcaact taaacagtgg 1380

tgattgacaa tggagagaat aggattgcag agctaattca gaggcagaat taactggact 1440

taataattga ttagccatgg ggatgaacaa gacaggaaga aatcaaggat catctggaag 1500

tttctattct agagactgat tattgtgtca tttagtgaga gaggtaatac caaaaaaaaa 1560

gtgcatgttt gaaaggagag ataactaatt cagttcggga tatatggaat ttgaggtgtc 1620

tgttgggtat tgtagatgtc tagcaggtta gatggacata tggatgtgga gctccaagag 1680

aacagcctgg gtaaaatata tggatgtgaa agttctcaaa aaggatgaca aactaaactg 1740

gggtgttagc agtgagcaca ataggaaaga acagctttca aagatataaa aatttagaat 1800

agataggatc tagtgattga ttggaataga taggatctag tgattgattg gataaaggga 1860

atgtgaaaaa gagacaaaac ctcagatttt tggtttgaaa gacaaacctc acatttttgg 1920

tttgaaagac aaacctcaga ttttggtttg aaagactggg tgcgtgacaa ttccactatt 1980

ataagctaag ttatgttccc tcaaaattca tatgttgatg ccctagcccc cagtatttct 2040

atatttgact gtatttggtg atagggtttt taaagaggta attaaggtta agtgaagtca 2100

tatgaatggg accttatcca atataactga tgttcttata atgaaaggag attaggatac 2160

aggcatacat agagggagaa tgtcatgtga atatggagat ggtcatctac aagccaagga 2220

gagaggactc agaagaagcc aaccctgata acaccttgat cttggacttc tagcctccag 2280

agcagtaaga atataaattt ctgccattta agctactcag tctgtggtat tttgttatgg 2340

cagccctagc aaactaatat agccaccaac taaggttggg aatatgaggg gaaaaagtag 2400

gttaggagaa ggtgttgaaa gaagactttg atatagaaca tgttgagttt gagattcctg 2460

aggtatatcc aggtggagat gttcaaatgg ccttgaaatt aaaagatata actgggttac 2520

aggtacagat gtggaaatca ctgatatgta ggtgatattt tcaactgtag gaatgaatga 2580

tatcatctac gaagtctgtg taaagtgaga atgatcatca agaagagaac cataaggcat 2640

gtgaacatta aaaagcagac tgaggaagag gacccaagaa aggagaaaaa tcagagacag 2700

aagatgaaga gagagagctt agaattagtc tttattttat cctgttttca ttataattag 2760

ttaagtacat attgatagca ccatttaaag agctttaact tagtgaagga ggcaatatct 2820

tattcatttc tgaatctcaa acaccaagca cagtgcctga catttaatat ccctgcaata 2880

aatatttgtt gaaccaaatg gaatcgaaat taggacttca ttcccttctt tgttttgttt 2940

gcttatttct ggagtgtgca gcatgtattg ccactgatat tttaagttca tagtgaaaaa 3000

tcatttcaat gtgttctagg cttggtcccc ttccaaccca acccccagga caatagtttt 3060

cattataacc ttctgtcttt gccattctgg atggaattct gtgcacagaa gttatataca 3120

tatatgggta tatctatgta acaaatcgca gcacaggagt cccctgggct ccctcaggct 3180

ctggtatgac atatttgagc catataaatt cagcttctcc tctggcatct gttagccgac 3240

tcacttgcaa ctccacctca gcagtggtct ctcagtcctc tcaaagcaag gaaagagtac 3300

tgtgtgctga gagaccatgg caaagaatcc tccagagaat tgtgaagact gtcacattct 3360

aaatgtaagt tgattcatat tttttccctt ttgagcagaa gcatggtttc acagatttat 3420

catattgcaa agtgaacatt agaaagtgga atcaaaggtg actttgaatt atggcttgct 3480

tattgagatt tattgtgtat tgttttattc tctctgttct ttggttaggc agaagctttt 3540

aaatccaaga aaatatgtaa atcacttaag atttgtggac tggtgtttgg tatcctggcc 3600

ctaactctaa ttgtcctgtt ttgggggagc aagcacttct ggccggaggt acccaaaaaa 3660

gtaagtaaat acacatcata atctgatgct tctgttctga gtttgattga atttaattag 3720

taggcatata aattattctg aaaatgaaaa tcattaagtc tcttttgtgt ttatttttctg 3780

gtggtatgaa aaactgaatc aagatttttc tctccttcat ttgcctaatt attttggtaa 3840

gggaaaaaaa tttatcttct atattttggt tagtaatgat agggtgtaga attacccatt 3900

cttctataat gcttatattt aaggagttaa gactcctcct aggagcaaag tttttacctt 3960

aagaaaaaat acttttaatc tttctttcaa ttatctgggc agaaactgat ggaaaagtaa 4020

ttagaccacc acataaggag aaaccaccca aagtggcagc aaaggatttt cagggattat 4080

gtttctgact taatgcatct gcagtgagaa caggcagcaa gctgcttttc aggaggagca 4140

attctcatac ctgacttcac atttatttat gactttaaaa agaaactaga gcttttcagc 4200

tccttgccag ctgcaatcat cactgtccca ctgagatagc aatgccacaa ggtagaagca 4260

tgcgacacca agtcattaat cacagataac tcagaaagct tctgatactg gggtcatttg 4320

gtgaagcttg gctagccact gtagtttata tttctgtacc ttagtgttat tgttattcct 4380

gggctcatag gcaatgggca agtttagata taaaataaat aaataaaacc atagtcccaa 4440

ataaaaccat agtcccttca tgctcacaat ttccatttga aagataatca acaaatatac 4500

aaactacaag ttgtgaagat agttttctgg ctttgccatg tttttctgtt ggtgtgtgga 4560

ttgctgctga agttattgtt attactgttg ttaattggtt gtcataagat gctcatttct 4620

atttggtaac agatgtaggg agaaaaaagc acaaatccag gaaggaggat gatacaagaa 4680

aatgggttgg catggaccgt taggagccaa aggagcaata ggagcagaga gaaaaaaata 4740

tgggtcagtg gggagagaga ggcagagcag gaaaagcaga aaataagctg gctaagtgtc 4800

tgaaagataa cagaggtata aagcataacc acccttgggt tctccacctc tgctgtcata 4860

cccactttca tttagtccca aaggatatct tatcaggttt aagtaggagg gagagcaagc 4920

cagcctactt tgtttgcctt tttatttata ctttgttcca gaaaagattt caggcatcac 4980

ccaatgccta tagttagccc tatcccctac tcctcctccc cttactgctc cttgcttctg 5040

aactacttac cgttcaaatg tccttctaga gaagccacct gtgttcaggc tggcagggtg 5100

ggtaaagcaa tcgagatcag gaaggaaaca gcaaattaat aagagagcca aaatgcattc 5160

caatacaatt ctggtaccac ttagctaatt cccaaggtga tatgtcaccc cacaagttga 5220

accttggatc ctcttgaaaa ttatccatga gttaaagggc cataaatcct ttgttctcca 5280

cagccataat gtacttaact gggcatggtg gatagctagt ccttttgacc acccttagat 5340

gcttgcctaa ataaggggcc aaatggcctt tcttgtttta ttgttgtcat tggtactgca 5400

aatgctattc caaaagagta tatgtgtttt tataaattta cagtgctatt tgcctcaacg 5460

cagctataaa cgactaggtg gagaccttgc ctttttattg caaagtgtca aactcacttg 5520

cagcttcatg gtctcctctc tttccctatt atatcatctt tacatctcat tctagttctg 5580

tttttaatga ttataaggag caatcctatg atacctcaaa ccaaaataaa acttggtcag 5640

tgcatgcaca gaagaggtgg gatatgctgc tgaataaagt ttatcaaaaa tcaaacctct 5700

tgtgaagtaa tccattttgg gatgtcattt tcagcaaagg gactctcaga aatgaacaaa 5760

gctattatga agtaatccag ataaagacaa caggtgtccc ctagggcaag tttctctacc 5820

atgaacaaat attgttggtt ttccattcca gtggggagtt atgattcatt gtctttgata 5880

tttaagtcat ggccttcata agaccaggtt tggttcattg aaaactaaga aatcagctgg 5940

aaatgaggat gctttactta gtccactact accttcaggt cccagacaaa gatctcaaga 6000

gaccatctat gtttccaatg ggtacgccta tcaacaagcc tgccttagag cagacttaag 6060

aattctggga agagactagc agtcagatta gtcttggaag agttgcatag gaaagggaca 6120

caggcatctt acataaatta aagactctcc ccttaatggt atggcctagc aagaaacaac 6180

agcagattaa aatccaggct ctttgggctt tccttaacaa atctatgaga cttgttggga 6240

ctctgtggcc cttggtagag agagctgtgc atgcaactgt gcttattggt ggcagtgatt 6300

tgtttgcagt cctgggattt cctccaagct cttagtgttt gctctcctag cctccaccac 6360

gcctcctttt tattttgtct ttcaacatct gtgtataata aacagtattt atttaacatc 6420

ccacatctgg tgttgacatt tccatgccaa taagtccttc tcctcactcc cgctttggtg 6480

tgtacaaaag atccagattc ctgtatccaa attgagaaca agccatgctt gacctcctcc 6540

cctcacatac ctcaggtctg cgtgcacaca ggtgtgcaca cgcacagact cacacacaca 6600

cagacacaca cacacacaca cacagaccag agagagaatg aaatcaaaat aaagtgaaac 6660

aaggacgact aactctgggt gagctgctgc attttcttac aggtcccagg cccctgaccc 6720

acagccacat acaggagcca cctgagctaa atgaaagcca ggtccatcca gctctgaatt 6780

cagctttctt aaacaagcag aagagaccta ttcagtccca gcaattttga gaaagcatcg 6840

taatgaggct ccccttctag gcagtgactg ccctgggatc attgaagcag aatcaccaaa 6900

gaaacattct tgcttaaagc cccctgtatt acacatttta ctggaaacag tcagaagcca 6960

actcctcagc taattaatcc aacatgaagc aaacaaatca gaaaaatcct agcccttaaa 7020

ccaacagttt agaaacttag cacagacctt gtagcccatc cctgccaatc ctttgtggag 7080

cgaggcagaa gaatcataca tttaataaat gaaagagtgt ccttactgcc ttacatagga 7140

gctgcctggt gactttaggg ttgtttgaaa gccatcagga ccacagtaat caatagattc 7200

gtgaggatgt tccttcaact tctatatatt accagagcaa tgaggagtgg tagggataag 7260

cgggcacttt tctgccctcc tgccacacct cagttgcagg tgatcactaa aacagtctcc 7320

cccagcgttg tgagtccact atcattctag gttgaaaatc ctttgctcag aagaagcact 7380

ttcttgacga gtctgattgc aacattttgt gcgtgcctat ttcttggcac gcacaattgg 7440

gcattagccc aatattaact tttttcgctc ctctctaggc acacaaggaa gacaactttt 7500

ttctccaggt cccctgggct ttcctacacc acagtccggt gggtgaaatg tgggtctaga 7560

gaaacattaa tataccctca aggaaaaagt tgaaggagta gccggaatat ttttggcagg 7620

gagactgagc tttaccagat ggtgccatgt tggcattttc aaaaaaaaaa aaaaaaaaaa 7680

gacctctctt cattttcctc ttaaagcatt gtgagaccca ggtccaccct tgtacaactc 7740

tcactgggcc agctcttatt ctgggctact caagggcagc tcccagcaat tatgtggagc 7800

caggctgtgg ctgtgatggg catgttcttc ttagtcagtc tcttcattct cctaaggcag 7860

catttgctga acaacttata tgacaccagg atgaagagcc tgggccttca aagacagctg 7920

ttctggagtg gttgagagtg gaattggggc tgtggaggaa gtagcctgga tgatcaacag 7980

atagtggcag taggcgtagt ccagtgatca acctgacagg ttgatctgag caaatctcag 8040

cagactttgc taactatcta acattacaga tgaaaagggg taaaggatta aaataggaac 8100

tgagtcccta gagacactca gtgatttgcc caaaatcaca cagcttgtaa gtggtggggt 8160

taaggttaga actagaggct agtctgaaga ttctttccac tattcttcag ctgcatttaa 8220

gaatcaggga tggggataat gggtaggaga aaaacttact ggtcagctca cgtctgtaga 8280

taaaagaagt agctttttct tgttgaaaaa gaatgtctgt gctgcttatg ccaatatttt 8340

gcatattgcc aaatggtgat tgaagatgca gaactctaac tttgcataag gaacaatctg 8400

tctgctctgt gttggtaaag aaatacccag aggcccttta attccccgcg agttaggtta 8460

ataatagata tggtagagtg cattctcttc ccttccttcc agcctaacac ctgctagcca 8520

gctattctga gcatgtattt caccctttgga gagtgtacct cataacagac agtaagggat 8580

gaagatggag tgaagagcag ctagattaat cacaggcttc aaaaaacaaa cctaagccaa 8640

atgtcctcaa atccctgtca ggacaggcta aagcttcttg cccctaaaac agatccatca 8700

cactgaggtc tgagatagct ttacctttтc taaatagctt tcttttctga tctcagggac 8760

tgccacagag gcaagaaaac aaggaaaact gtgaaagagg gatgttaggg ctcatttgct 8820

gcctgaatgg aggatttaac tggactgaag acagaattcc tgggcaggga agaaaaatgg 8880

gttgtgtcct ctacatattt tgtcagtgtg gtcttcttac acttctcact gcctggactt 8940

aaccacatat ctgtctcatt tcacttctga ttcctattcc actccattgg cttccttgtt 9000

ccattccagt gttctggttc ttatagttaa ggccagcctt atgtgtgggt ggagtaagag 9060

cccatttgag tggtgctgga aagatgccct tgcccccaaa tatctcccat tcctcgtgcc 9120

actgtgccta ttagattggt gcaaaagtaa ctgtggtttt agccattgaa agtaatggca 9180

aagtaataaa gtaataaaat ccactattag tacttcattc ctctgagtga agtattgcac 9240

tcaaattttt aaacagacaa atattcatca ctaatatcta agtgtgaaat catgggtcag 9300

attgatttct tctttgtcat aaggtacctg aagcaccacg agattcctag atgtctagac 9360

aggagacaaa gggacataaa tgggaagaga aaggggaagg gctatgaggg tatagaatga 9420

acagaatatg aaaatagagc agaaaaggaa agctagaagc accacaaatt atctctctct 9480

attttccaat tgggcagggg tgtactcctc atcatattct gtccttgctc tatcctccct 9540

ggaataaagc tgagttggag gcagctgaga tgagctactt gtagcccact cacacctctc 9600

ttcaggccag tcatcctgtc tctgctttag cttgctcctg gagaggacaa agccatttat 9660

tctttcatct ccagaaagac cttcaacccc tgagccagga agcttagtag agagcaaagt 9720

acttgttatg tcaaatggca tgagactcta cgtcttcata cgggccagtg aagcttgcat 9780

agccattacc caagcaatgt gctccacagg gacccaggcg agacagtgtg tctggatttg 9840

ctcataccca tctccaaaca gtccaaagtg ggccaccagc atcatcttcc tagacgaaga 9900

acagtctgtt ccaaaaggac acctttgata gtggaattct gtctgggtga atatatagct 9960

ccccactcct acccccctacc aaggattcac acataaacga agtcattctg ttcccactct 10020

agtaacttat cattactaac ataaagccaa caagttggtt aatgaacaca atcagaatgg 10080

acccccttctc ccctaccaaa aaaactcttt gcaggaatgt tgaaaataat caagatctgc 10140

tgagcttggc agtttccaat agttttcatc tagcaaggtg tccaagtaaa gtgcattgcc 10200

aacagtgcag gaaaactcat tccacagaag cacactgaaa ttctaaatgt caccagtagg 10260

ttccatagtc ttctaccagt gattgaaaat acagcctgta ttttctcttt cgcccaggta 10320

ggcagtgttc cagcatcccc tagtctctaa cctcattttt ccatacaata atccatcagt 10380

ggtgctactg atgaaataaa caccccaccg cagcccccaa tggctagtgg catcaacatt 10440

tgtctcaaaa atggatagta tttctaacaa aattcccccg ttgggagagc tgacaggact 10500

ggatctgaag tggatgcctc tcttttctac ttaggtcaga cagactgaga gctcaatttg 10560

ttgccacagg tagaagtttc tcttctcaag ctggctgcag agctggaact gaggacatat 10620

ttattgctag aagcaaataa caaacctaaa ggaatttgct acctatcaca aaccaggcca 10680

gacttcctga gttaactctg cagttgccag agagttctaa gtattgtttt tctttcactt 10740

ttaaaatggt ttttttaaaa gaccttgcca atactcaaaa atgtctcttt aggaagataa 10800

ggatgtgcca ttggtgggga caacttattc aacaaatact taggttccaa ccatgtccaa 10860

agcactgtag ggctaccaag aggtaaatga cattgtctct tgtttttctc ttcagcagct 10920

acattagact tgtcaccctt gaagtaacac ttgcttctac aaatggttcc aatttgtttt 10980

tctctagttg agttggtgaa tatttcacca ggttaaacta acattcagtc aatggcatca 11040

gcagagacta aaaaatgacc caacctggtg taagttctct ataaggctgc cactgggaac 11100

cgtatcatta gaaagcaatg tgggctctga agaagacttg ctttttcagg cttttagttt 11160

cccagatttt tttcataagg aaaaatgaac atggagaaac tcgataaaca aaggtctcct 11220

aatgacagag ataagataga cttccaaaag gcatcatcag gtctggcctg gtttgtgaca 11280

gataacaaat tcctttatgc ttatttgctt ctacctataa atatgtccta aaactataag 11340

ttatgggctt actcctaaac tgcagtgcca tatgggactg gtcctcttgg tctttagaaa 11400

tacaattgcc tcttttagag tgcaggagaa aatctggggt tttgttttag ccactcaaga 11460

aagacctgat gagcagtctt tcttcttggg aggctccatg ttacttaatg caggacactt 11520

ggctcagttg gcttaattcc atatacacaa aagaatccgt agaacttgaa aacaactgaa 11580

tatggcttgg gaaacagaac ccaggaataa actgggtggg gttggttgcc gtataaagag 11640

aaggctaagg gaaggcaata aatgtttcct ggagtttaag gggcaaaagg aagcaaacag 11700

agcagggtag ggctctaata ttggcaagga gaaagtcttc agattctaaa aggaatagac 11760

tcaaggctgt tgtggcaata ttaaagccct ggataaataa ggagacaaaa aacacttaac 11820

cattttaagt gtattcaaat ctctttctaa gccctaatga atttcacggg tgaggggcgg 11880

ggcagtaagg gggggtcgt gattacttag cacttggaaa gggaagtggt agcttaaaga 11940

agcagcatag gttttaggtg atccctcagc ttaacacaag gggaaaatac tttatagget 12000

ggtttgcaaa ctatcatttg ctgtttagtc aaggctgcca agaaaactgt tggaattctt 12060

agtaattagc tcagcagctt ggcttgaatt caaaatacca gctctgaagg gatccccata 12120

tcagctccag cgccaaaaag cacctcactt tagggacaca ctgagtatct tagagattgt 12180

tttcctctgt tctcccaggc ctatgacatg gagcacactt tctacagcaa tggagagaag 12240

aagaagattt acatggaaat tgatcctgtg accagaactg aaatattcag aagcggaaat 12300

ggcactgatg aaacattgga agtgcacgac tttaaaaacg taagttggat gttttcctcc 12360

taaggctttc cacttaaaat attagagcag ttgagtcaag ttaaaaatag cctccatcta 12420

aaatttgaat tcaaatggaa agctagcctc ctctagcttt ccatttgaat tcaaattttg 12480

ctagtctggg accccaacag atcatggaac ttactaatag tggctctttt gggaagcttt 12540

attgttgttt tgtctgtttt atagggatac actggcatct acttcgtggg tcttcaaaaa 12600

tgttttatca aaactcagat taaagtgatt cctgaatttt ctgaaccaga agaggaaata 12660

gatgaggtat gtaagaagaa taattgtggt ggcaaaagac atcatttatt ggatgctagc 12720

tatgtgccaa acattgtact aaatgcttta cttttctttg attctccaac aaccttataa 12780

atagatacta ttgttatcat cattttcaga tcaggaaact agaactcaga gccaaaaagt 12840

gacttgctca aggccacaca gctagtgagt agcagattgg ggtcttgaac ttaggccttt 12900

cagaaaactc aagtgcataa ttttaggtgc aatgctatac tcttataccc gggcacccat 12960

taacattatg ttgattttct ctgaaaccaa agatttctag taagagggat gagccagaga 13020

aacaaaatca atcttgcctc agataaagcg tttttaaaaa agaaagttcc aaagccacct 13080

ggggttcttg ccacttttgt cacacttcag tattcaccta ttttaaagat tgtggttcat 13140

tcaccttacc attttatgct gggctgctaa tggaataatt cacatttaac ctcagtagga 13200

aaggcaaaag caaagaaggt gaacactgca atgggataca aattctctga aaacaccctg 13260

ctagaccaaa gtaaggcaac ctagggtgct ggtgaaatca ctgaatggcc ctttaaaacc 13320

taatgggaga tggatgttat tgcctaaaat acaacgatct gtaaggccct tgaggctctt 13380

tattgttccc tgatattttg tctataatta cagtatggtt ttattagtgg gaggagaagg 13440

gggacctcag ttaaggtcat attgtgacag catatcccca ctatagcaag aaaaaacaaa 13500

cattggtttg gtttggttta gtgcttagtt gccaaatcta aagtggaatt atggcaatgg 13560

tagaatgatc tttcttcttc ttcttcttct tctttgtatt tttagtagag acggggtttc 13620

accatagcca ggatggtctc aatctcctga cctcatgatc cgcccgccta ggcctcccaa 13680

agtgctggga ttacaggcat cagccaccgc gcccagccat ggcagagtga tctttctata 13740

aactataaaa acctctttct ttctttcttt ctttcttctc tctctccctg ttctttttca 13800

cttttctccc acccagctcc ctccatcttg attctcacca acccatttat cttttccttt 13860

acctttttct ttaccttttt ctttatttgg gggcggtggg tggggaggac tatttttctt 13920

tcccttaccc aagtcttaga gctgtcattg gctcccactt tactctttca tgctactggt 13980

ctcctcgggg ccagaggagt tacagaccct tgttactaga atttgcaaga aaaaaaagaa 14040

cgttctccca aagatgatgc tctgctccca aggtctacgg tagtgacaaa gtttggagac 14100

ttttttaaa gtaagcataa gttagatgca tctgaggaat agatcgacgt tcaggatatt 14160

aggaatgact atagaaggca aaggtaagat aaaaggaaaa tcaataggat taaaaagcta 14220

ctatacacaa taacatatgc tgattacata cagaaaataa tgtaagcatt tgagtaggcc 14280

actcagcaat gagaacctga agtgatataa taggatttat ttttttgaaa gattaaggca 14340

aaggaggcgg ggcacggtgg ctcacgccta taatcccagc cctttgggag gccgaagagg 14400

gggatcactg gaggtcagga gttcgagacc actctggcca acatggtgaa aacgcatctc 14460

tactaaaaca aaacaaaaca aaacaaaatt agccgggcat ggtggcgatg tgcttgtaat 14520

cccagctact gaagaggatg aggtgggagg atcacttgaa cccgggagtg ggaggtttca 14580

gtaagccgag tttgcgccac tgcactccag cctgggtgac agagcgggac tcagtgtcaa 14640

aaaataaata aataaacaaa aaataaaggc aaaggctcat aatgtctcat tttggcccta 14700

atttagtctt cagttaattt caccaaatat tttaaataac gcaaattcat ttcatgatag 14760

taagaacact gtaccttcaa gtaaatcttg ttttttatat tcacgccaca gatttttttt 14820

taatgagaca aatttgtggt cctgaatgac aaaaaccttc ttctgtaatt ccctccacac 14880

tgtatccatg gcatgtgtat tttaaacatg cagagcaggt atataaatag ccctgatttt 14940

gctaacaaag ctagtcagca taccaacctc attttataca attcctttca atgcatctca 15000

aggtcgttac agactcaagg acaaaggtgc taagacttta gaattgtaaa tgattccggg 15060

agaaattagt aagaactagg tatacaacac aattacctca atactagaaa ctgaagagaa 15120

agtggcatcc ttcacaatat aaaaaaagct gcttaaagga aaaacactct gtgagataaa 15180

ctctaagcat ttacaatgaa aggaatgcaa tagtaggaag gccagtgaag tcataaactg 15240

gagacttgcc tttaaaacat aaaaggaaat cttttgtact ataactagct accatcaaca 15300

ggcaaatctt gtaaaaccaa aatattagat tacattgtta ttgaattgat ctcaccagtt 15360

tggattatag gagttttgtc agcttcgggt caaatcctgc ttgccaccca agatgcaaac 15420

ctaaaataag gcaggctaat aaaatgtatt tataaaaaag aaaacaaatg taaatcactt 15480

agggttccag aagctcaagt taaaacatac acagacccag ggcaccctga aaataacgaa 15540

gccccttttct gactttgagt gaaatggtaa cagagaaggc tcctacttta gtaagcagtc 15600

ttgctgttag acccaggaag gaggaaggat cagttgttaa ctcccccaag ggcacccact 15660

gaatgcaagg cacagagcta gttcaccagc agccctagga ttaaaattgt tacactcaaa 15720

atatattgct ttccagcatc ccataatgtc tcacttgttc tactaatttc tctgccaatc 15780

cctaccctaa gctactttct tctttcagaa tgaagaaatt accacaactt tctttgaaca 15840

gtcagtgatt tgggtcccag cagaaaagcc tattgaaaac cgagattttc ttaaaaattc 15900

caaaattctg gagatttgtg ataacgtgac catgtattgg atcaatccca ctctaatatc 15960

aggtatgaca ttcttatcct catcctcctc ctattttcta agacagtaac aggttaatgt 16020

tcccagcatt tagaggctac atttctagga aaacaaatga tcggtttata tcttcttgga 16080

tgagtctata tatgctgact gccagtgagg agggaatttt tcacagccaa gttataaatc 16140

agataaatta atgttgacag gcttttaaat atcccagcaa aggtaacacc accctaatgt 16200

gactagtgat agcaaaaaac taatctgtct ctctttgcac atctatattc attgcagtcg 16260

gtatgccaga ctattgctca aaccactgtg ggtccaaccc aagaatagag gagagaaggg 16320

caaaaacaat gctctgctat caagaaggat tccctatgtc aaatggtggc attttcctta 16380

gaagatagga caatcactgc aatacaaatt tctttgttgc tgttttagga attactataa 16440

aaatagttga gggcttgatg tccagagtag aagagtccag ccctcaactt ccagtcagcc 16500

catagagaag gtaaaggccc ctggatgtct tagtctgttt agctaacatg gaatttggcc 16560

tccctcattc tgcacagcat agattactct attacccaag agaggttatt ctcaccaatg 16620

taagcatctg tgggtactct ttagagatat taaaaaggtg actaaatgca gtgtggtatc 16680

ctggactgga tgctggaaca gaaaaaaaca ttagtagaaa tactaatgaa atctgaataa 16740

agtctgtagt ttagttaata acattgtacc aattctaact tctgattttt gatcattgta 16800

ccatggttat gtgaggtaac attaggggaa accgggtgaa gaatatgtga actctgtact 16860

atttctgcaa ctcttccata aatctaagat tatttctaaa cagaaagttt aaaaaaaatt 16920

taagacaaga aacataaggg aagtgggagg acagaggatg tcaaatcaat tcaggcaacc 16980

aagaaatatt tattgagcat ctacatgggt aaaatcatca ttgtaggaga tacagaagtc 17040

ccttgtcctc agtttttaat ttagtgtgct attaagtcaa agcagtaaat gtttgatgcc 17100

aaagaagcaa agaagtttgt gtatgtagaa atgatcaggc ttgttttgga aactactcta 17160

aaatatcagc tgaaggagct tggaaaagtg gtccatgcct aaatgatgcc agttagaaga 17220

cacaatgaaa atctctatcc ccagctgcta ttctcccaag cacttctttc aggaacattt 17280

gcaaagcagt tgcatcacaa ctttgcattt attatcttag tttctgagtt acaagacttt 17340

gaggaggagg gagaagatct tcactttcct gccaacgaaa aaaaagggat tgaacaaaat 17400

gaacagtggg tggtccctca agtgaaagta gagaagaccc gtcacgccag acaagcaagt 17460

gaggaagaac ttccaataaa tgactatgtg agttatgttt atgtaaactc ttgatagagg 17520

aggaaaaaag agccctcaca aaacttacta cccctcagat cacaatcatt ggaaggagaa 17580

cgatcatggc tttaacaaaa aaaaaaaaat gtttaagaaa cttggaatct ccaggaggct 17640

cattcaaaat caatctcgat ctctgtagat atatctccat gatcaggcta aggagcttaa 17700

tttctcacaa gcctccaact catacagtct tctagttctc ccctgcttgg tgagaactct 17760

tgctagtgaa ccttcctctc ccacccccaa cactggcttc ttcttctttc agactgaaaa 17820

tggaatagaa tttgatccca tgctggatga gagaggttat tgttgtattt actgccgtcg 17880

aggcaaccgc tattgcggcc gcgtctgtga acctttacta ggctactacc catatccata 17940

ctgctaccaa ggaggacgag tcatctgtcg tgtcatcatg ccttgtaact ggtgggtggc 18000

ccgcatgctg gggagggtct aataggaggt ttgagctcaa atgcttaaac tgctggcaac 18060

atataataaa tgcatgctat tcaatgaatt tctgcctatg aggcatctgg cccctggtag 18120

ccagctctcc agaattactt gtaggtaatt cctctcttca tgttctaata aacttctaca 18180

ttatcaccaa cagcctgatt gctgctgaac acactgggta agtgtttgct gagaatattc 18240

acagcaactg aaatggagca gacccgtcaa acatagatga ctatgctgag gctcagaaaa 18300

caatgcccca aaatgaaagc ctcagaagca aaagtttttc tctgaccttc tcctacctgc 18360

cagtctctca gtcccattct cccccgaggc tagccataaa agctagaatc cctcttcctc 18420

aaggcgtgtc atataaatca aaaccccttt tttccaaagc cagccttaaa cctaaaaata 18480

tcactctgat ttttcctctg cttttctgtg taaaagctgg ccataaggaa attatctgac 18540

ctaccttgtt tgactgtagg tcataagact ccctttccag agaggatcca acacaacagg 18600

tttggaagta gaaaagggca aagcataatg atacctaaat atcatatttt actctgctgt 18660

attaagacag tggaaataag gagctgatga taaaatgtct ccataactga tgcattaaaa 18720

acatatcctt ttgtgcatag ccccagtgtt taattatatg aaagatatat tgtgaacttc 18780

cttaagtatc aattaagctg ctggacaggc agtgagtcat agcaggccaa cactaaaata 18840

ttccaaatgt gcagactcag caaaggagaa agctaaacta gcaaaaaggg tctcaggctc 18900

tcctcctata aacaggaact caacttagga caaattcaac attccatagc cttcacatct 18960

<210> 11
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer 2244-23

<400> 11
cacgaagtag atgccagtgt atcc 24

<210> 12
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 12
gtgtacttcc aatgtttcat cagtgc 26

<210> 13
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 13
ccagttacaa ggcatgatga cacg 24

<210> 14
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 14
cgtcctcctt ggtagcagta tgg 23

<210> 15
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 15
ccatcctaat acgactcact tagggc 30

<210> 16
<211> 23

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 16
actcactata gggctcgagc ggc 23

<210> 17
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 17
gtcagtgatt tgggtcccag cag 23

<210> 18
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 18
cgtgaccatg tattggatca atccc 25

<210> 19
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 19
atggcaaaga atcctccaga gaac 24

<210> 20
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 20
ctattagact ctcccaagca tgcg 24

<210> 21
<211> 11

<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: peptide

<400> 21
Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg
1 5 10